The Journal of Experimental Education

A periodical report of scientific investigations relating to child development, curriculum, learning, teaching, supervision, measurements, statistics, and experimental techniques.

Vol. XXIX | June 1961 | No. 4

CONTENTS

Page

A Study of the Relationship of Scores on the MMPI to Success in Teaching as Indicated by Supervisory Ratings
Carroll Edward Flanagan 329

A Comparison of Teachers' Response Patterns on the MMPI with Response Patterns of Selected Non-Teacher Groups
Elmer Christian Rieck 355

A Comparison of the Readability of Vocational Agriculture Reference Books with the Reading Ability of the Students Using Them
Ralph R. Bentley and R. Edward Galloway 373

The Relationship of Certain Letter Form Variants in the Handwriting of Female Subjects to Their Education, I.Q. and Age
Lawrence Epstein, Huntington Hartford, Irving Tumarkin 385

A Study of Pathological Doubting as a Response Set *Velma I. Rust* 393

Geometric Aid in Teaching Rectilinear Correlation *John Gaito* 401

INDEX

$7.50 A YEAR | PUBLISHED QUARTERLY | $1.90 A COPY

Published by Dembar Publications, Inc.,
Madison 3, Wisconsin.
Second class postage paid at Madison, Wisconsin

EDITORIAL BOARD

A. S. Barr, Chairman, Professor of Education, University of Wisconsin, Madison 6, Wis.

Jacob O. Bach, Head, Education Research Bureau, Southern Illinois University, Carbondale, Illinois. Editorially responsible for materials on university field research.

R. O. Collier, Associate Professor, College of Education, University of Minnesota, Minneapolis. Editorially responsible for materials on measurements, statistics, and methods of experimental research, published each March.

Arthur T. Jersild, Professor of Education, Teachers College, Columbia University, New York City. Editorially responsible for materials on child welfare, guidance, and development, published each December.

H. H. Remmers, Professor of Educational Psychology, Director Division Educational Reference, Purdue University, Lafayette, Indiana. Editorially responsible for materials on learning, teaching and supervision, published each September.

J. Wayne Wrightstone, Director, Bureau of Educational Research, Board of Education of the City of New York, Brooklyn, New York, 110 Livingston Street, Brooklyn, New York. Editorially responsible for materials on curriculum construction, published each June.

CONTRIBUTING EDITORS

Emmett A. Betts, Director, Betts Reading Clinic, Haverford, Pennsylvania.

Leo J. Brueckner, Professor Emeritus, University of Minnesota, Minneapolis, Minnesota.

Oscar K. Buros, Professor of Education, Rutgers, The State University, New Brunswick, New Jersey.

Guy T. Buswell, Executive Secretary, American Educational Research Association, 1201 Sixteenth St., N.W., Washington 6, D. C.

Harold D. Carter, Professor of Education, University of California, Berkeley 4, California.

Leslie L. Chisholm, Associate Professor of Education, State College of Washington, Pullman, Washington.

Herbert S. Conrad, Technical Consultant, College Entrance Examination Board, Princeton, New Jersey.

Stephen M. Corey, Professor of Education, Teachers College, Columbia University, New York, New York.

Robert A. Davis, Professor of Educational Research, George Peabody College for Teachers, Nashville, Tennessee.

Hari R. Douglass, Director Emeritus of College of Education, University of Colorado, Boulder, Colorado.

Harold A. Edgerton, Director, Occupational Opportunities Service, Professor of Psychology, Ohio State University, Columbus 10, Ohio.

John C. Flanagan, Professor of Psychology, University of Pittsburgh, Pennsylvania.

Carter V. Good, Dean, College of Education and Home Economics, Teachers College, University of Cincinnati, Cincinnati 21, Ohio.

Robert W. B. Jackson, Professor of Educational Research; Assistant Director, Department of Educational Research, Ontario College of Education, University of Toronto, Toronto, Canada.

D. Welty Lefever, Professor of Education, University of Southern California, Los Angeles, California.

Edward A. Lincoln, Consulting Psychologist, Halifax, Massachusetts.

Irving Lorge, Professor of Education, Executive Officer, Institute of Psychological Research, Teachers College, Columbia University, New York 27, New York.

A. R. Mead, Director Emeritus of Educational Research, University of Florida, Gainesville, Florida.

T. E. Newland, Professor of Education, University of Illinois, Urbana, Illinois.

C. W. Odell, Professor Emeritus of Education, University of Illinois, Urbana, Illinois.

Willard C. Olson, Dean, School of Education, University of Michigan, Ann Arbor, Michigan.

Valworth R. Plumb, Chairman, Division of Education and Psychology, University of Minnesota (Branch), Duluth, Minnesota.

S. L. Pressey, Professor Emeritus of Educational Psychology, Ohio State University, Columbus, Ohio.

William Reitz, Professor of Education, College of Education Examiner, Wayne University, Detroit 2, Michigan.

Henry D. Rinsland, Professor Emeritus of Education, The University of Oklahoma, Norman, Oklahoma.

Robert T. Rock, Jr., Professor of Psychology, Head of Dept. of Psychology, Graduate School, Fordham University, New York City.

Phillip J. Rulon, Professor Emeritus of Education, Harvard Graduate School of Education, Cambridge 38, Massachusetts.

John Schmid, Research and Evaluation Psychologist, Lackland Air Force Base, San Antonio, Texas.

Louis G. Schmidt, Associate Professor of Education, School of Education, Indiana University, Bloomington, Indiana.

David Segel, Director, Lincoln Guidance Research Project, Public Schools, Albuquerque, New Mexico.

Helen Thompson, Assistant Clinical Professor in Clinical Psychology, New York University Bellevue Medical Center, 120 East 75th Street, New York 21.

Robert L. Thorndike, Professor of Education, Chairman of Department of Psychological Foundations and Services, Teachers College, Columbia University, New York City.

Herbert A. Toops, Professor of Psychology, Ohio State University, Columbus, Ohio.

Maurice E. Troyer, Vice President, Japan International Christian University, Tokyo, Japan.

Helen M. Walker, Professor of Education, Teachers College, Columbia University, New York City.

Guy M. Wilson, Professor Emeritus of Education, Boston University, 33 Pine Street, Wellesley Hills, Massachusetts.

Paul A. Witty, Professor of Education, Director of Psycho-Educational Clinic, School of Education, Northwestern University, Evanston, Illinois.

Ernest R. Wood, Dean, Rocky Mountain College, Billings, Montana.

D. A. Worcester, Visiting Professor of Education, University of Wisconsin, Madison, Wisconsin.

DEMOCRAT PRINTING COMPANY
MADISON, WISCONSIN

The Journal of Experimental Education

PUBLISHED QUARTERLY

Copyright 1961 by

DEMBAR PUBLICATIONS, INC.
Madison 3, Wisconsin

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 4, June 1961)

A STUDY OF THE RELATIONSHIP OF SCORES ON THE MMPI TO SUCCESS IN TEACHING AS INDICATED BY SUPERVISORY RATINGS

CARROLL EDWARD FLANAGAN*
Wisconsin State College
Whitewater, Wisconsin

SECTION I

The Problem and Its Background

THE PURPOSE of this investigation is to study the relationship between certain aspects of personality and teacher efficiency. More specifically, the study is of the relationship between scores and profile patterns on the Minnesota Multiphasic Personality Inventory and success in teaching as indicated by principals' and superintendents' ratings.

There is considerable evidence from previous research to indicate that teacher ratings based on pupil gains and supervisory ratings of teachers are not highly correlated. Rostker[1]** and Lins[2], for example, each concluded that pupil gain and supervisory ratings are not related to a greater degree than can be attributed to chance. Lins stated, "Whatever is measured by each of the criteria, at least it appears evident that these criteria do not measure the same aspects of teaching efficiency."[3] LaDuke[4] summarized, "Ratings of teachers do not agree with the criterion of pupil gain."

Supervisory ratings of teachers, then, must pertain to something other than efficiency in promoting pupil gain. They probably pertain to the personal characteristics of teachers and effectiveness in discipline. An examination of the verbal statements relative to the strengths and weaknesses of the teachers rated by the principals and supervisors responding to the inquiry would seem to indicate that, except for discipline, the primary concern of the raters was with the personal characteristics of teachers.

The value systems used by principals and supervisors in rating teachers differ to a considerable degree. This, which would tend to reduce correlations to zero, is a possible reason for the low correlations between the supervisory rating and the pupil gain criteria. It would seem that any relationships which have been found between supervisory ratings and other data exist in spite of the diversity in supervisors' value systems.

Regardless of the meaning of supervisory ratings, teachers are usually judged by their supervisors and administrators. In general, these evaluations are made upon the basis of general impressions made of the teacher in many personal contacts, in observing classroom performance, and upon reports from others, rather than from a single objective measure of teaching efficiency. The impressions may depend to a large extent upon the teacher's personality, but the evaluation, whatever its basis, becomes the official evaluation of the particular school system in which the teacher is employed. Regardless of what else the teacher may have done, right or wrong, if he pleases the administration he is rated highly and consequently is considered to be a good teacher.

It can be said, also, that in spite of the inadequacies of supervisory ratings, the local administrator or supervisor should be in a better position to judge the abilities of the teacher than persons who must base their judgments on limited observation. In any case, it seems that supervisory ratings are very important when considering a teacher's success.

As to the aspect of teaching here chosen for study, a teacher's personality is apt, according to the thinking and researches of many persons, to influence his success in teaching. Various qualities affect his classroom effectiveness, his enthusiasm, his pupil contacts, his discipline, his contacts with fellow teachers and with the citizens of his community, or other aspects of his life as a teacher. Certainly, extreme abnormalities in personality determine a teacher's effectiveness in guiding pupils in the learning-teaching

* The author wishes to express his thanks to Professor A. S. Barr, who served as his major professor, and to Professors L. E. Drake and T. A. Ringness, both of whom served on the author's thesis committee.

**All footnotes will be found at end of article.

situation. The study here reported purports to explore the influence of certain personality abnormalities as measured by supervisory ratings.

It is expected that this study will discover such relationships as may exist between teacher effectiveness as measured by supervisory ratings and scores on the MMPI.

SECTION II

The Procedures Used in the Investigation

Design of the Study

The study here reported attempts to compare the ratings of teachers who were rated outstanding, above average, average, below average, or poor by their principals or superintendents, with scores on the MMPI.

The Minnesota Multiphasic Personality Inventory

The MMPI is a psychometric instrument developed at the University of Minnesota by S. R. Hathaway, Professor and Director of the Division of Clinical Psychology at the University of Minnesota, and J. C. McKinley (deceased), formerly head of the Department of Neuropsychiatry at the University of Minnesota. The Inventory was designed to provide scores on all of the more important clinical phases of personality. Although it was standardized on 1500 normal individuals and 220 psychopathic patients in the University of Minnesota Hospital and was designed primarily as a clinical instrument, it has been used very extensively in a great variety of experiments of a non-clinical sort. Albert Ellis[5], in the Fifth Mental Measurements Yearbook[6], says, "It can confidently be stated that in the whole history of modern psychology there has been no other personality inventory on which so much theoretical and practical work has been done." The Fifth Mental Measurements Yearbook lists 495 studies about or with the MMPI. Accordingly, including those listed in previous yearbooks, there has been a total of 779 studies through 1958, relating to this particular instrument. Most of the early research on the Inventory was done in a hospital setting, but later research was conducted with normal persons and in relation to various situations. Drake and Oetting[7] state that "...almost all the published work has reported studies of the individual scales. While information about the individual scales may be useful in some instances, the behavior that the counselor is trying to understand or predict is usually too complex to be measured by a single scale." Accordingly, attention has been shifted to patterns of scales as well as individual scales.

The group form of the MMPI consists of 566 statements covering a wide range of subject matter, from physical condition of the individual being tested to his morale and attitude[8]. In administering the test, the subject is asked to respond to all statements which are in the first person, as True, False, or Don't Know. An IBM machine-scoring answer sheet has been adapted for this Inventory by Professor L. E. Drake of the University of Wisconsin Student Counseling Center, which has greatly facilitated the scoring operations.

The Inventory scales, of which there are ten plus four correction scores, were developed serially from 1942 to 1951. This probably accounts for the fact that most of the earlier research with the MMPI has been with individual scales rather than profiles.

Data-Gathering Procedures and Subjects of the Study

The MMPI was administered to entering freshmen during freshman orientation week at the University of Wisconsin. The scoring was done by the University of Wisconsin Student Counseling Center. Personality profiles for the subjects of this study were obtained from these score cards.

A list of all students who received the University Teachers' Certificate along with the teaching fields of these students was secured from the files of the Office of the Dean of the University of Wisconsin School of Education. The teaching positions held by the subjects were obtained from the files of the University Teachers' Placement Office, the University Alumni Office files, and from the official directories of the public schools of the state of Wisconsin.

Letters were sent to all known principals and superintendents of teachers in the study asking their cooperation in rating the teachers involved. The rating blank asked for comments regarding elements of strength in the teacher's work, the elements of weakness in the teacher's work, general comments, and an overall rating on a five-point scale.

Tabulations of the ratings, comments, and patterns of personality scores were then made and relationships among the various data sought.

The subjects for this study consisted of all the students enrolled at the University of Wisconsin during the fall semesters of 1949, 1950, and 1951, who received the University Teachers' Certificate in 1953, 1954, 1955, or 1956, and who have taught in the public schools of Wisconsin. It was found that a total of 431 students from the freshman classes of 1949, 1950, and 1951 had taken the MMPI, had graduated in the years 1953-56, and had received the University Teachers' Certificate. Due to the Korean situation and other matters, many of the males who had originally enrolled did

not graduate and many more who did graduate did not go into teaching. Marriage and decisions not to teach removed many of the females from the list of teachers. For others, no record of post-graduation activity could be found. In some cases, due to the death of a principal or superintendent, or to his moving to an unknown location, it was impossible to get a supervisory rating on a particular teacher. In a few cases the administrators and supervisors failed to supply the ratings. The number of teachers for whom complete data were secured was 167, of which only 20 were males.

Although this is not intended to be a sampling study, there appears to be no reason to believe that the people here studied were greatly different from other people who graduate from the University of Wisconsin School of Education. Since this is a very limited population, users of the data in this study may be interested to know more about the group. Considering the group of 167 as a unit, this population was different from the usual in that only twenty were males. For this reason, and because high or low scores on many of the MMPI scales mean different things for men and women, the female population of 147 was studied separately. Of this group, nine were music teachers; twelve, teachers of art; thirteen, speech correction; fifteen, physical education and recreation; thirty-three, home economics; six, business education; fifteen, elementary; eight, social studies; one, mathematics and science; and thirty-five, English, foreign language, speech or library. All had grade-point averages of 2.5 or better at the beginning of their junior year. In general, they appeared to be a typical University of Wisconsin School of Education group. Assuming there has not been much change in the basic characteristics of University of Wisconsin students, the interested reader is referred to Lins[9] study for detailed information concerning a typical university group of women students.

Whenever it was possible, which was in more than half the cases, two or more ratings were secured on each teacher. When these ratings did not agree, they were averaged and the average rating used. When ratings differed, they differed by only one rank.

Of the 147 women, ten were rated outstanding or 1; eight had a mean rating of 1.5; sixty-one were rated above average or 2; fourteen had a mean rating of 2.5; forty-two were rated average or 3; three had an average rating of 3.5; seven teachers were rated below average or 4; and two were rated poor or 5. Either most of the teachers in this study were above average or supervisors and principals have a tendency to overrate their teachers. Because the frequencies in some of the categories were rather small, they were combined as follows: eighteen teachers in the 1 or 1.5 rating group, seventy-five cases in the 2 or 2.5 rating category, forty-five in 3 or 3.5, and nine cases in the 4 or 5 rating group.

Coding the Data

The scores on the MMPI had been coded according to the method of Hathaway.[10] The scales were assigned numbers as follows:

0	Si	- Social Introversion-Extroversion
1	Hs	- Hypochondriasis
2	D	- Depression
3	Hy	- Hysteria
4	Pd	- Psychopathic Deviate
5	Mf	- Masculinity-Femininity
6	Pa	- Paranoid
7	Pt	- Psychasthenia
8	Sc	- Schizophrenia
9	Ma	- Hypomania
X	None	

Raw scores had first been converted to standard T scores. (T scores have a mean of 50 and a standard deviation of 10). In the coding, the number of the scale with the highest T score was written first, then the second highest, and so on down. No score which fell within half a standard deviation on either side of the mean was recorded. This means that only scores of 55 or above were coded high. After the high scores had been listed according to the number of their scales, a dash separated the low scored scales. These were listed by scale number with the lowest scored scale first, the next lowest second, and so on. No score above 45 was coded low. Tied scores were underlined. Thus, if a person's profile code is 3987-645, it means that the highest score was made on scale 3, the score on scale 9 was equal to that on 3 or differed by only one point, the next highest score was on scale 8, and the next on scale 7. Scores of 55 or above were made on all of these scales. The lowest score was made on scale 6, the next lowest on scale 4, and the next on scale 5. The scores on these three scales were all below 46. Scores on the remaining scales (0, 1, and 2) were within the normal or average range.

This system of coding has been recommended because it expresses both intensity and pattern and at the same time effects a compromise between too much detail on scale elevation and equally undesirable detail as to profile shape[11]. Du Mas[12], in discussing problems of profile analysis concludes, "If prediction or diagnosis is to be maximally good, both intensity and pattern must be operative in a profile."

Methods of Analyzing the Data

There are many theories of mental structure, organization, and techniques for the validation of data-gathering devices with reference to these.

The conventional statistical techniques for test validation appear to rest upon the assumption of the additive nature of the constituents of human abilities and personalities. This study does not make this assumption and, accordingly, a different technique of validation has been employed.

It is thought by many researchers in personality that patterns of personality traits, the ways in which various phases go together, are the important determiners of personality manifestations. In some cases one scale reinforces another, while in others one scale compensates for or offsets another. For certain scales extreme scores seem to be the only important ones. The most promising recent research, especially that of S. R. Hathaway and L. E. Drake, has been conducted with reference to these ideas.

In the present study, comparisons were made between supervisory ratings and highest ranking scales, combinations of high ranking scales, and combinations of high scales and low scales. It is hypothesized that personality profiles on the MMPI are different for highly rated teachers than for teachers who are rated lower. McCall[13] found a relatively close relationship between teacher personality and pupils' comprehensive growth as measured in his investigation.

Various procedures were employed in analyzing the data. Among these was the chi-square. When a chi-square large enough to be encouraging was found, the data were examined more closely and other tests were made to see if a relationship between a particular rating and a scale or profile pattern might be more firmly established. Since in most cases the chi-square statistic was used to test whether having a particular scale or combination coded in a certain manner or not having it coded thus was statistically significant in regard to supervisory ratings, the expected frequency of each cell becomes a matter of interest.

Inasmuch as there are forty cells in a 4×10 table which was employed in the overall chi-square analysis, the expected frequencies for 147 cases would naturally not be large. In many instances, they fell below the minimum of five or ten frequently thought necessary for such analyses.

Under the null hypothesis it is assumed that having a particular personality trait has no relationship to supervisory ratings. For example, as shown in Table I, twenty-four female teachers had scale 3 (Hy) coded highest of all the scales. Under the null hypothesis, one would expect these twenty-four cases to be distributed among the supervisory ratings proportionately to the way in which the total 147 cases are distributed among the supervisory ratings. Seventy-five of the 147 teachers were rated in the second category. Then the expected frequency for the cell designating scale 3 coded highest and supervisory rating of 2-2.5 would be that part of twenty-four that seventy-five is of 147. This is 12.24.

In this use of chi-square, the data are put into various categories (e.g., the dichotomy of having a scale coded high or not high on one hand and the four-fold categorization of supervisory ratings on the other). These categories should not be interpreted as points on a scale.

Chi-square is a measure of the amount that observed frequencies differ from expected frequencies in the various categories. In this study most of the observed frequencies were about what one would expect. This is indicated by the chi-square values, only a few of which were large enough to be statistically significant.

Tate and Clelland[14] say, "The test (chi-square) is the best available for enumeration data from a manifold population. It gives good results even when the expected frequency in one or more categories is as low as one, provided that the majority are about five or more." Not all authors agree with Tate and Clelland's statement regarding the use of chi-square and particularly the minimum number required for expected frequencies. Since the number in each category is rather small in some cases (after the data are broken down into sub-categories) chi-squares have been figured in several instances where the statistical significance of the result must be seriously questioned. Many authors do[15,16,17], however, indicate that the heretofore commonly accepted minimum expected frequency of five per cell is too high, especially when there are two or more degrees of freedom. In this study there are three degrees of freedom for most of the chi-square tests.

In addition to the chi-square tests, the percentages of occurrence of the various patterns of high and low scores were calculated for each of the four rating categories. It was thought that such data would enable one to make certain qualitative analyses of profile patterns in relation to the supervisory rating. The following profile patterns were compared:

1. The frequency with which each scale of the three highest scales occurred with every other scale coded high (above T = 54).
2. The frequency with which each scale of the three highest scales appeared with each of the two lowest scales. This was done for each of the four rating categories.
3. The frequency with which each scale coded anywhere high (above T = 54) occurred with each scale coded anywhere low (below T = 46). This was done for the highest rated group.

The data obtained in these comparisons are shown in the tables reproduced in Section III.

TABLE I

FREQUENCY OF SUPERVISORY RATINGS OF FEMALE TEACHERS WHO HAD VARIOUS MMPI SCALES CODED HIGHEST OR TIED FOR HIGHEST

Scale Coded Highest or Tied for Highest		Supervisory Rating				Chi Square
		1 - 1.5	2 - 2.5	3 - 3.5	4 - 4	
N =	147	18	75	45	9	D. F. = 3
0	11	2 (1.35)*	7 (5.61)	0 (3.37)	2 (0.67)	7.177
1	3	1	1	1	0	
2	3	1	1	1	0	
3	24	7 (2.94)	11 (12.24)	5 (7.35)	1 (1.47)	7.934
4	11	2 (1.35)	4 (5.61)	4 (3.37)	1 (0.67)	1.142
5	18	2 (2.20)	12 (9.18)	4 (5.51)	0 (1.10)	2.733
6	18	1 (2.20)	10 (9.18)	6 (5.51)	1 (1.10	.893
7	30	1 (3.67)	16 (15.31)	11 (9.18)	2 (1.84)	2.954
8	8	0	5	2	1	
9	30	4 (3.67)	16 (15.31)	9 (9.18)	1 (1.84)	.560
X	4	1	0	3	0	

*Number in parentheses following observed frequency is expected frequency for that cell.

SECTION III

The Analysis of Data

The purpose of this study, as stated in Section I, is to explore the relationship between scores and profile patterns on the MMPI and success in teaching as indicated by supervisory ratings. The study is designed to discover whether or not there are differences in the personality patterns of successful and unsuccessful teachers.

In order to carry out the stated purposes of this study, it was necessary to submit the data to certain statistical analyses which will be explained. The data in this study are nonparametric in nature and, consequently, any statistical measures employed should be of the nonparametric type. The methods of securing the supervisory ratings and of grouping them into four categories were explained in Section II, as was the method of coding the scores from the MMPI.

The Female Teachers

The following discussion is in reference to the group of 147 female teachers only.

The frequency with which each scale was coded highest or tied for highest for each of the four supervisory rating categories was first tabulated (Table I). Because of the low frequencies in some of the categories certain planned techniques had to be abandoned. However, chi-squares were figured for cases which had a total N of 11 or more. Whenever the chi-square value was large enough to suggest some relationship, closer examination of the data were made. In most cases chi-square values were calculated using a method advocated by Lindquist[18] which is less time consuming than the traditional method, but which is the algebraic equivalent of the traditional.[19]

Only one group—the group with scale 3 (Hy) coded highest—seemed worthy of further investigation. A chi-square of 7.934 with three degrees of freedom was found for this group, indicating that having scale 3 (Hy) coded highest might discriminate among the female teachers in the various rating categories. With three degrees of freedom, a chi-square of 7.934 is significant beyond the 5 percent level.

For the group with scale 3 coded highest, the expected frequencies for each rating, under the null hypothesis of no difference, were 2.94, 12.24, 7.35, and 1.47 for ratings of 1 - 1.5, 2 - 2.5, 3 - 3.5, and 4 - 5, respectively. A comparison of the expected frequencies with the observed frequencies indicated that the greatest discrepancy was for the 1 - 1.5 rating. It was then hypothesized that being rated high, i.e., 1 or 1.5 was related to having scale 3 coded highest of all the scales. The data were arranged in a 2×2 contingency table, Yates' correction[20] for small expected frequencies applied, and chi-square calculated. The value found was 5.871 (see Table IX). With one degree of freedom, this value of chi-square is significant beyond the 2 percent level. Examination of the data showed the relationship to be positive.

For these data at least, there seems to be a potential relationship for female teachers between the occurrence of scale 3 (Hy) being coded highest and a supervisory rating of outstanding. Much larger groups of teachers including many who have been rated outstanding are needed to establish this relationship firmly.

Continued exploration of various relationships among the data tabulated as indicated above failed to reveal additional statistical significant relationships. However, an examination of the relative frequencies in the various categories did bring out some interesting relationships which will be pointed out. Although only eleven cases out of the total 147 female teachers had scale 0 (Si) coded highest, only two of these received supervisory ratings of average or less. It would appear that an absence of scale 0 (Si) coded highest may be associated with average or below average supervisory ratings.

However, when one considers the small number (eleven) in the group, it is apparent that this may be a false conclusion. When the group is enlarged to include all those who had scale 0 (Si) coded among the three highest scales (see Table II), the number is increased to thirty-one and the number receiving ratings of average or below is increased to thirteen, which is more than the expected frequency. This would seem to dispute the conclusion of relationship between a lack of scale 0 (Si) being coded high and average or below average supervisory ratings.

The above is an example of a fact that has been obvious throughout the analysis of the data in this study, namely, that for studies of this nature very large groups, preferably in the thousands rather than the hundreds, are highly desirable. Whereas normally a much smaller group is ample for a good investigation, studies of this kind necessitate a group large enough to provide sufficient cases in each of the sub-groups after the data have been broken down.

A similar examination of the group with scale 7 (Pt) coded highest indicates a possible slight relationship between scale 7 (Pt) coded highest and average or below supervisory ratings. When the relationship between scale 7 (Pt) being coded among the three highest and supervisory ratings is investigated, support is gained for this possibility, since the chi-square for this relationship is 5.254 (about the 16 percent level). Although the evidence is not statistically significant at the commonly designated 5 percent level, analysis of the data for this study seems to give some indication that high codings of scale 7 (Pt) are not

TABLE II

FREQUENCY OF SUPERVISORY RATINGS OF FEMALE TEACHERS WHO HAD VARIOUS MMPI SCALES CODED AMONG THE THREE HIGHEST

Scale Coded Among the Three Highest		Supervisory Rating				Chi-Square
		1 - 1.5	2 - 2.5	3 - 3.5	4 - 5	
N =	147	18	75	45	9	D. F. = 3
0	31	4 (3.80)*	14 (15.82)	11 (9.49)	2 (1.90)	
1	14	2 (1.71)	8 (7.14)	4 (4.29)	0 (0.86)	1.135
2	16	1 (1.96)	9 (8.16)	4 (4.90)	2 (0.98)	2.178
3	54	9 (6.61)	25 (27.55)	15 (16.53)	5 (3.31)	.774
4	45	5 (5.51)	25 (22.96)	13 (13.78)	2 (2.76)	.691
5	29	3 (3.55)	20 (14.80)	6 (8.88)	0 (1.78)	5.761
6	45	6 (5.51)	22 (22.96)	15 (13.78)	2 (2.76)	.576
7	62	4 (7.59)	32 (31.63)	20 (18.98)	6 (3.80)	5.254
8	42	3 (5.14)	24 (21.43)	12 (12.86)	3 (2.57)	1.862
9	48	6 (5.88)	25 (24.49)	15 (14.69)	2 (2.94)	.474
X	4	1	0	3	0	

*Number in parentheses following observed frequency is expected frequency for that cell.

TABLE III

FREQUENCY OF SUPERVISORY RATINGS OF FEMALE TEACHERS WHO HAD VARIOUS MMPI SCALES CODED LOWEST OR TIED FOR LOWEST

Scale Coded Lowest or Tied for Lowest		Supervisory Rating				Chi-Square
		1 - 1.5	2 - 2.5	3 - 3.5	4 - 5	
N =	147	18	75	45	9	D. F. = 3
0	21	3 (2.57)*	13 (10.71)	3 (6.43)	2 (1.29)	3.248
1	5	0	3	2	0	
2	16	3 (1.96)	10 (8.16)	3 (4.90)	0 (0.98)	3.009
3	7	1	5	1	0	
4	4	1	2	0	1	
5	41	4 (5.02)	17 (20.92)	17 (12.55)	3 (2.51)	3.625
6	7	0	3	4	0	
7	13	2 (1.59)	3 (6.63)	8 (3.98)	0 (0.80)	7.626
8	3	1	1	1	0	
9	11	1 (1.35)	7 (5.61)	1 (3.37)	2 (0.67)	5.091
X	19	2 (2.33)	11 (9.69)	5 (5.82)	1 (1.16)	.413

*Number in parentheses following observed frequency is expected frequency for that cell.

associated with high supervisory ratings for women teachers, but that low codings of scale 7 are associated with average or below supervisory ratings for women teachers (see Tables III, IV, VII and IX). The code type 7 - 5 (high coding of seven with low coding of 5) was a common one for women teachers in both the above average and average or below rated groups, the evidence again favoring the lower rated teachers. Superior teachers did not, in general, make their highest MMPI scores on scale 7 (see Tables I and II).

Further examination of the data revealed a possible meaningful relationship between scale 5 (Mf) coded among the three highest and supervisory ratings. The chi-square value indicating the relationship was 5.761 (about the 13 percent level). This was high enough to merit further investigation (see Table II).

The number in the group which had scale 5 (Mf) coded among the three highest was 29. The expected frequencies for the four supervisory ratings from 1 to 5 were, respectively, 3.55, 14.80, 8.88, and 1.78. The largest discrepancy between expected and observed frequencies was for the group with ratings of 2 or 2.5. When the chi-square test was applied to the re-grouped data to test whether scale 5 being coded among the three highest might be related to a 2 or 2.5 rating, a value of 4.65 was obtained. If Yates' correction for small expected frequencies is applied, the calculated chi-square value is 3.80 (see Table IX). With one degree of freedom, 3.80 just misses being significant at the 5 percent level.

The data for this same group were then arranged in a 2×2 contingency table to test whether having scale 5 (Mf) coded among the three highest scores might be related to ratings anywhere above average, that is, from 1 through 2.5. With this arrangement, Yates' correction was unnecessary since the expected frequencies in each cell were above ten. This time a chi-square of 4.00 was found (see Table VII). This is significant beyond the 5 percent level, indicating that for the group of women teachers in this study, a high coding of scale 5 is positively related to above average or outstanding supervisory ratings. There is some indication that the rating is more apt to be 2 or 2.5 than 1 or 1.5.

There is some indication that a low coding of scale 5 (Mf) for the women teachers is associated with average or below average supervisory ratings (see Tables III, IV, VII, and VIII).

A comparison of the frequencies of various combinations of high and low coded scales for the "above" and "below" rated teachers did not indicate any statistically significant differences. The number in most categories was too small to be dependable (see Table VIII). An examination of the data, however, shows that for every code type with 5 low (0 - 5, 1 - 5, 2 - 5, 3 - 5, etc.,) the expected frequency is larger than the observed frequency for the average or below average group.

Since scale 5 coded low has been found to be either a suppressor or an intensifier factor with high codings of other scales[21], chi-square tests were run to see if a low coding of scale 5 with a high coding of each of the other scales differentiated between supervisory ratings above average and those below that. For the group of women teachers studied, no sigificant differences in supervisory ratings were found within the group with a particular scale coded hign in respect to having scale 5 coded low or not coded low. Data for this analysis can be found in Table VI.

A low coding of scale 2 (D) was associated with the group of women teachers studied with above average or outstanding supervisory ratings (see Tables III, IV, and VII). When the supervisory ratings were divided into a dichotomy, a chi-square value at the 5 percent level was calculated (see Table VII). Drake[22] says that a low coding of scale 2 for women is frequently related to "socially outgoing characteristics".

The proportion of the number in each supervisory rating category who had the various scales coded highest or among the highest was also calculated and recorded as a percent (see Appendix B, Tables A and B).* Tabulations of other combinations were also made (see Appendix B, Tables C, D, and E). While some of the relationships shown in these tables may be of some interest, none of these supplementary tables indicate anything of statistical significance.

The Male Teachers

The number of male teachers in the group studied was so small (twenty) that little could be done of a statistical nature. When the available data were broken down into the necessary subgroups, the number in each group was usually zero or one and seldom more than two. Descriptive statements concerning this group of male teachers will be made as far as seems appropriate.

Two of the twenty men in the group received supervisory ratings of 1.5; thirteen of the group were rated 2; one was rated 2.5, and four were rated 3. Thus, sixteen of the twenty men in the group were rated above average or outstanding by their supervisors or principals. The other four were rated average. With numbers this small, any statements of relationships between supervisory ratings and MMPI code types could not be

*All references to Appendices will be found in original thesis on file, Library, University of Wisconsin, Madison, Wisconsin.

TABLE IV

FREQUENCY OF SUPERVISORY RATINGS OF FEMALE TEACHERS WHO HAD VARIOUS MMPI SCALES CODED AMONG THE TWO LOWEST

Scale Coded Among the Two Lowest		Supervisory Rating				Chi-Square
		1 - 1.5	2 - 2.5	3 - 3.5	4 - 5	
N =	147	18	75	45	9	D. F. = 3
0	36	4 (4.41)	20 (18.37)	9 (11.02)	3 (2.20)	1.113
1	22	3 (2.69)	12 (11.22)	6 (6.73)	1 (1.35)	.303
2	32	5 (3.92)	20 (16.33)	6 (9.80)	1 (1.96)	3.919
3	14	1 (1.71)	8 (7.14)	5 (4.29)	0 (0.86)	1.522
4	16	2 (1.96)	7 (8.16)	6 (4.90)	1 (0.98)	.466
5	50	5 (6.12)	22 (25.51)	18 (15.31)	5 (3.06)	3.623
6	15	2 (1.84)	7 (7.65)	6 (4.59)	0 (0.92)	1.582
7	27	5 (3.31)	11 (13.78)	11 (8.27)	0 (1.65)	4.882
8	10	2 (1.22)	4 (5.10)	4 (3.06)	0 (0.61)	1.764
9	23	4 (2.82)	13 (11.73)	4 (7.04)	2 (1.41)	2.603
X	19	2 (2.33)	11 (9.69)	5 (5.82)	1 (1.16)	.413

*Number in parentheses following observed frequency is expected frequency for that cell.

TABLE V

FREQUENCY OF SUPERVISORY RATINGS OF FEMALE TEACHERS WHO HAD VARIOUS PAIRS OF MMPI SCALES CODED AMONG THE THREE HIGHEST

Pairs of Scales Among the Three Highest		Supervisory Rating				Chi-Square
		1 - 1.5	2 - 2.5	3 - 3.5	4 - 5	
N = 147		18	75	45	9	D. F. = 3
01	4	1	1	2	0	
02	5	1	1	2	1	
03	7	1	4	1	1	
04	4	1	3	0	0	
05	3	1	1	1	0	
06	8	1	3	4	0	
07	21	2 (2.57)*	9 (10.71)	8 (6.43)	2 (1.29)	1.379
08	11	0 (1.35)	7 (5.61)	4 (3.37)	0 (0.67)	2.683
09	2	0	0	2	0	
0X	0	0	0	0	0	
12	1	0	1	0	0	
13	8	2	4	2	0	
14	3	0	1	2	0	
15	1	0	1	0	0	
16	2	0	2	0	0	
17	4	0	3	1	0	
18	1	0	1	0	0	
19	5	1	3	1	0	
IX	0	0	0	0	0	
23	4	0	3	1	0	
24	5	1	4	0	0	
25	1	0	1	0	0	
26	2	0	1	1	0	
27	10	0	5	3	2	
28	5	0	3	1	1	
29	1	0	0	1	0	
2X	0	0	0	0	0	
34	15	2 (1.84)	6 (7.65)	5 (4.59)	2 (0.92)	1.873
35	9	1	6	2	0	

TABLE V (Continued)

Pairs of Scales Among the Three Highest		Supervisory Rating				Chi-Square
N = 147		1 - 1.5	2 - 2.5	3 - 3.5	4 - 5	D. F. = 3
		18	75	45	9	
36	19	2 (2.33)	10 (9.69)	6 (5.82)	1 (1.16)	.097
37	11	0 (1.35)	6 (5.61)	2 (3.37)	3 (0.67)	10.772
38	6	0	3	2	1	
39	15	3 (1.84)	6 (7.65)	5 (4.59)	1 (0.92)	1.267
3X	5	1	2	2	0	
45	5	0	4	1	0	
46	11	2 (1.35)	5 (5.61)	4 (3.37)	0 (0.67)	1.271
47	20	1 (2.45)	11 (10.20)	7 (6.12)	1 (1.22)	1.257
48	10	0	6	4	0	
49	16	2 (1.96)	10 (8.16)	4 (4.90)	0 (0.98)	1.749
4X	0	0	0	0	0	
56	6	0	4	2	0	
57	4	1	1	2	0	
58	2	0	2	0	0	
59	11	0 (1.35)	9 (5.61)	2 (3.37)	0 (0.67)	.499
5X	5	1	3	1	0	
67	19	2 (2.33)	9 (9.69)	8 (5.82)	0 (1.16)	2.387
68	11	2 (1.35)	5 (5.61)	4 (3.37)	0 (0.67)	1.271
69	8	2	2	3	1	
6X	2	0	1	0	1	
78	31	2 (3.80)	17 (15.82)	9 (9.49)	3 (1.90)	2.032
79	15	1 (1.84)	9 (7.65)	4 (4.59)	1 (0.92)	.782
7X	0	0	0	0	0	
89	13	2 (1.59)	7 (6.63)	3 (3.98)	1 (0.80)	.459
8X	0	0	0	0	0	
9X	4	0	1	3	0	

*Number in parentheses following observed frequency is expected frequency for that cell.

TABLE VI

FREQUENCIES OF PRESENCE OR ABSENCE OF LOW CODING OF SCALE 5(Mf) WITH HIGH CODING OF OTHER SCALES FOR FEMALE TEACHERS WITH ABOVE AVERAGE OR SUPERIOR AND AVERAGE OR BELOW SUPERVISORY RATINGS

Code Type* High - Low	Supervisory Rating 1 - 2.5 N = 93	Supervisory Rating 3 - 5 N = 54	Chi-Square D. F. = 1
0 - 5	6	7	
0 - Not 5	12	6	1.304
3 - 5	8	7	
3 - Not 5	26	13	.826
4 - 5	10	8	
4 - Not 5	20	7	1.667
6 - 5	12	7	
6 - Not 5	16	10	.001
7 - 5	16	15	
7 - Not 5	20	11	1.060
8 - 5	10	8	
8 - Not 5	17	7	1.046
9 - 5	7	7	
9 - Not 5	24	10	1.838

*High coded scale was among the three highest; 5, when coded low was among the two lowest.

TABLE VII

COMPARISON OF FREQUENCY OF OCCURRENCE OF CERTAIN CODE TYPES FOR FEMALE TEACHERS WHO WERE RATED ABOVE AVERAGE OR SUPERIOR WITH THOSE WHO WERE RATED AVERAGE OR BELOW AVERAGE

Code Type	Supervisory Rating 1 - 2.5 N = 93	Supervisory Rating 3 - 5 N = 54	Chi-Square D. F. = 1
0 highest	9 (6.96)*	2 (4.04)	1.760
3	18 (15.18)	6 (8.82)	1.704
4	6 (6.96)	5 (4.04)	.556
5	14 (11.39)	4 (6.61)	1.869
6	11 (11.39)	7 (6.61)	
7	17 (18.98)	13 (11.02)	.707
9	20 (18.98)	13 (11.02)	.188
0 among 3 highest	18 (19.61)	13 (11.39)	.460
1	10 (8.86)	4 (5.14)	.442
2	10 (10.12)	6 (5.88)	
3	34 (34.16)	20 (19.84)	
4	30 (28.47)	15 (16.53)	.323
5	23 (18.35)	6 (10.65)	4.002
6	28 (28.47)	17 (16.53)	
7	36 (39.22)	26 (22.78)	1.245
8	27 (26.57)	15 (15.43)	
9	31 (30.37)	17 (17.63)	
0 lowest	16 (13.28)	5 (7.72)	1.761
2	13 (10.12)	3 (5.88)	2.499
5	21 (25.94)	20 (15.06)	3.550
7	5 (8.22)	8 (4.78)	3.775**
9	8 (6.96)	3 (4.04)	.458
2 among 2 lowest	25 (20.25)	7 (11.75)	3.886
5	27 (31.63)	23 (18.37)	2.799
7	16 (17.09)	11 (9.91)	.228
9	17 (14.55)	6 (8.45)	1.330
07 among 3 highest	11 (13.28)	10 (7.72)	1.249
34 among 3 highest	8 (9.49)	7 (5.51)	.709
37	6 (6.96)	5 (4.04)	.389
49	12 (10.12)	4 (5.88)	1.066
67	11 (12.02)	8 (6.98)	.271
78	19 (19.62)	12 (11.38)	.066
89	9 (8.23)	4 (4.77)	.215

* Numbers following observed frequencies are expected frequencies.

**Chi-square for 7 coded lowest was reduced to 2.715 when Yates' correction was applied.

TABLE VIII

COMPARISON OF FEMALE TEACHERS RATED ABOVE AVERAGE OR SUPERIOR WITH THOSE RATED AVERAGE OR BELOW AS TO COMBINATIONS OF HIGH AND LOW CODED SCALES

Code Type High with Low	Above Frequency Observed	(Expected)	Below Frequency Observed	(Expected)	Chi-Square D. F. = 1
0 - 1	4	(3.80)	2	(2.20)	
0 - 2	3		1		
0 - 3	5	(5.06)	3	(2.94)	
0 - 4	1		1		
0 - 5	6	(8.23)	7	(4.77)	1.8064
0 - 6	2		2		
0 - 7	0		1		
0 - 8	0		0		
0 - 9	7	(6.96)	4	(4.04)	
0 - X	3	(2.53)	1	(1.47)	
1 - 0	1		0		
1 - 2	3		1		
1 - 3	0		1		
1 - 4	2		0		
1 - 5	2	(2.53)	2	(1.47)	
1 - 6	1		0		
1 - 7	1		0		
1 - 8	0		0		
1 - 9	3		1		
1 - X	2		1		
2 - 0	2		0		
2 - 1	0		1		
2 - 3	0		1		
2 - 4	0		0		
2 - 5	3	(4.43)	4	(2.57)	
2 - 6	1		0		
2 - 7	0		0		
2 - 8	1		0		
2 - 9	1		2		
2 - X	3		0		
3 - 0	11	(12.65)	9	(7.35)	.6779
3 - 1	4	(3.16)	1	(1.84)	
3 - 2	7	(5.06)	1	(2.94)	
3 - 4	3		2		
3 - 5	8	(9.49)	7	(5.51)	.7092
3 - 6	2		0		
3 - 7	7	(8.23)	6	(4.77)	.5495
3 - 8	5	(4.43)	2	(2.57)	
3 - 9	7	(6.33)	3	(3.67)	
3 - X	5	(5.06)	3	(2.94)	
4 - 0	7	(6.96)	4	(4.04)	
4 - 1	1		0		
4 - 2	5	(3.80)	1	(2.20)	
4 - 3	4	(3.16)	1	(1.84)	
4 - 5	10	(11.39)	8	(6.61)	.5264
4 - 6	2		1		
4 - 7	5		1		
4 - 8	0		2		
4 - 9	3		1		
4 - X	6	(5.06)	2	(2.94)	
5 - 0	8	(6.33)	2	(3.67)	.2881
5 - 1	7	(5.06)	1	(2.94)	
5 - 2	7	(6.96)	4	(4.04)	
5 - 3	2		0		
5 - 4	4		2		
5 - 6	5		2		
5 - 7	8	(6.96)	3	(4.04)	.4561

TABLE VIII (Continued)

Code Type High with Low	Above Frequency Observed	(Expected)	Below Frequency Observed	(Expected)	Chi-Square D. F. = 1
5 - 8	2		0		
5 - 9	5		0		
5 - X	1		0		
6 - 0	7	(5.69)	2	(3.31)	
6 - 1	4		1		
6 - 2	7	(5.06)	1	(2.94)	
6 - 3	2		0		
6 - 4	4		4		
6 - 5	12	(12.02)	7	(6.98)	
6 - 7	5		3		
6 - 8	2		1		
6 - 9	8	(6.96)	3	(4.04)	.4561
6 - X	2		2		
7 - 0	3		2		
7 - 1	3		2		
7 - 2	8	(6.33)	2	(3.67)	1.2881
7 - 3	6		3		
7 - 4	2		0		
7 - 5	16	(19.61)	15	(11.39)	2.2922
7 - 6	1		3		
7 - 8	0		1		
7 - 9	6	(5.69)	3	(3.31)	
7 - X	6	(6.33)	4	(3.67)	
8 - 0	4		1		
8 - 1	2		0		
8 - 2	6	(4.43)	1	(2.57)	
8 - 3	3		2		
8 - 4	0		0		
8 - 5	10	(11.39)	8	(6.61)	.5264
8 - 6	1		1		
8 - 7	1		0		
8 - 9	2		1		
8 - X	5	(5.06)	3	(2.94)	
9 - 0	13	(11.39)	5	(6.61)	.5264
9 - 1	7	(6.96)	4	(4.04)	
9 - 2	10	(8.86)	4	(5.14)	.4415
9 - 3	1		1		
9 - 4	2		1		
9 - 5	7	(8.86)	7	(5.14)	1.1755
9 - 6	2		1		
9 - 7	7	(8.86)	7	(5.14)	1.1755
9 - 8	1		1		
9 - X	6	(5.06)	2	(2.94)	
X - 0	1		0		
X - 1	0		1		
X - 2	1		0		
X - 3	0		1		
X - 4	0		2		
X - 5	0		0		
X - 6	0		2		
X - 7	1		0		
X - 8	0		1		
X - 9	0		0		

Note: Number rated above average - 93 (63.27%)
Number rated average or below - 54 (36.73%)

High coded scale was among the three highest.
Low coded scale was among the two lowest.

TABLE IX

FREQUENCY RELATIONSHIPS AMONG MISCELLANEOUS CODE TYPES AND SUPERVISORY RATING CATEGORIES FOR FEMALE TEACHERS

Code Type	Rating		Chi-Square D. F. = 1
	1 - 1.5	Below 1.5	
3 highest	7	17	
3 not highest	11	112	7.643 5.871 with Yates' Correction
7 highest	1	29	
7 not highest	17	100	2.786
3 among 3 highest	9	45	
3 not among highest	9	84	1.556
7 among 3 highest	4	58	
7 not among highest	14	71	3.349 2.479 with Yates' Correction
8 among 3 highest	3	39	
8 not among highest	15	90	1.424
47 among 3 highest	1	19	
47 not among highest	17	110	1.131
78 among 3 highest	2	29	
78 not among highest	16	100	1.227
7 among 2 lowest	5	22	
7 not among lowest	13	107	1.211
	2 - 2.5	Not 2 - 2.5	
5 among 3 highest	20	9	
5 not among highest	55	63	4.648 3.80 with Yates' Correction
	3 - 3.5	Not 3 - 3.5	
5 lowest	17	24	
5 not lowest	28	78	3.152
	4 - 5	Not 4 - 5	
7 among 3 highest	6	56	
7 not among highest	3	82	2.349
7 among 2 lowest	0	27	
7 not among lowest	9	111	2.157

TABLE X

FREQUENCY OF SUPERVISORY RATINGS OF MALE TEACHERS WHO HAD VARIOUS MMPI SCALES CODED HIGHEST OR TIED FOR HIGHEST

Scale Coded Highest or Tied for Highest	Supervisory Rating				Total
	1.5 N = 2	2. N = 13	2.5 N = 1	3. N = 4	N = 20
0	0	0	0	0	0
1	0	1	0	0	1
2	0	1	0	0	1
3	0	2	0	0	2
4	0	0	0	1	1
5	2	5	0	2	9
6	1	2	0	0	3
7	1	2	0	1	4
8	0	0	0	0	0
9	0	1	1	0	2

TABLE XI

FREQUENCY OF SUPERVISORY RATINGS OF MALE TEACHERS WHO HAD VARIOUS MMPI SCALES CODED AMONG THE THREE HIGHEST

Scale Coded Among Three Highest	Supervisory Rating				Total
	1.5	2.	2.5	3.	
0	0	2	0	0	2
1	0	2	1	0	3
2	0	3	0	2	5
3	1	5	1	0	7
4	0	2	0	2	4
5	2	7	2	2	13
6	1	4	0	0	5
7	1	3	0	2	6
8	1	4	1	1	7
9	1	1	1	1	4
X	0	0	0	0	0

TABLE XII

FREQUENCY OF SUPERVISORY RATINGS OF MALE TEACHERS WHO HAD VARIOUS MMPI SCALES CODED LOWEST OR TIED FOR LOWEST

Scale Coded Lowest or Tied for Lowest	Supervisory Rating				Total
	1.5	2	2.5	3	
0	1	2	0	0	3
1	0	1	0	1	2
2	0	0	0	1	1
3	0	1	0	0	1
4	0	1	0	0	1
5	0	0	0	1	1
6	0	1	0	0	1
7	0	2	0	0	2
8	0	2	0	0	2
9	0	1	0	0	1
X	1	2	1	1	5

TABLE XIII

FREQUENCY OF SUPERVISORY RATINGS OF MALE TEACHERS WHO HAD VARIOUS MMPI SCALES CODED AMONG THE TWO LOWEST

Scale Coded Among the Two Lowest	Supervisory Rating				Total
	1.5	2	2.5	3	
0	1	5	0	1	7
1	0	5	0	1	6
2	0	1	0	0	1
3	0	1	0	0	1
4	1	3	0	0	4
5	0	1	0	1	2
6	0	0	0	0	0
7	0	4	0	0	4
8	0	5	0	0	5
9	0	2	0	0	2
X	1	2	1	1	5

TABLE XIV

FREQUENCY OF SUPERVISORY RATINGS OF MALE TEACHERS WHO HAD VARIOUS PAIRS OF MMPI SCALES CODED AMONG THE THREE HIGHEST

Pairs of Scales Coded Among the Three Highest	Supervisory Rating				Total
	1.5	2	2.5	3	
01	0	0	0	0	0
02	0	0	0	0	0
03	0	0	0	0	0
04	0	0	0	0	0
05	0	1	0	0	1
06	0	0	0	0	0
07	0	1	0	0	1
08	0	2	0	0	2
09	0	0	0	0	0
0X	0	0	0	0	0
12	0	0	0	0	0
13	0	0	0	0	0
14	0	0	0	0	0
15	0	3	0	0	3
16	0	0	0	0	0
17	0	1	0	0	1
18	0	2	0	0	2
19	0	0	0	0	0
1X	0	0	0	0	0
23	0	2	0	0	2
24	0	0	0	1	1
25	0	2	0	1	3
26	0	2	0	0	2
27	0	0	0	1	1
28	0	0	0	1	1
29	0	0	0	0	0
2X	0	0	0	0	0

TABLE XIV (Continued)

Pairs of Scales Coded Among the Three Highest	Supervisory Rating				Total
	1.5	2	2.5	3	
34	0	1	0	0	1
35	1	2	1	0	4
36	1	1	0	0	2
37	0	0	0	0	0
38	0	0	0	0	0
39	1	1	1	0	0
3X	0	0	0	0	0
45	0	0	0	1	1
46	0	1	0	0	1
47	0	1	0	1	2
48	0	0	0	0	0
49	0	0	0	1	1
4X	0	0	0	0	0
56	1	2	0	0	3
57	1	1	0	0	2
58	1	4	0	0	5
59	1	0	0	0	1
5X	0	0	0	0	0
67	0	1	0	0	1
68	0	1	0	0	1
69	1	0	0	0	1
6X	0	0	0	0	0
78	1	1	0	1	3
79	0	0	0	1	1
7X	0	0	0	0	0
89	0	0	0	0	0
8X	0	0	0	0	0
9X	0	0	0	0	0

depended upon. The addition of one or two more cases could change the picture entirely in many of the categories.

It can be noted that nine of the twenty men had scale 5 (Mf) coded highest of all the scales (see Table X), and thirteen of the twenty had scale 5 coded among the three highest (Table XI). Only two men had a low coding of scale 5 (see Table XIII). Drake[23] says, "A high coding of this scale among men is particularly frequent in an educational setting."

No other scale was coded highest frequently enough to merit special attention. Scale 3 (Hy) was coded among the three highest seven times, all for above average teachers. Scale 7 (Pt) was coded among the three highest six times, two of these being for men who were rated average teachers. Scale 8 (Sc) was among the three highest scales seven times, distributed among the rating categories.

No pair of scales was coded among the highest three scales frequently enough to warrant any generalizations concerning it (see Table XIV). Pairs 15, 25, 39, 56, and 78 appeared three times each. Pair 35 was coded among the three highest four times and 58 was coded among the three highest five times. Four of these times were for teachers with supervisory ratings of 2 and the other was for a teacher with a rating of 1.5.

Five of the twenty men in the group had no scales coded low (more than one standard deviation below the mean). These five appeared in all four of the supervisory rating categories (see Table XII). Three of the group had scale 0 (Si) coded lowest and two each had scales 1 (Hs), 7 (Pt), and 8 (Sc) coded lowest.

Six of the men had scale 0 (Si) coded among the two lowest scales. A low coding of scale 0 for men is "probably indicative of an adequate social adjustment."[24] Five each had scales 1 (Hs) and 8 (Sc) coded among the two lowest. Four men each had scales 4 (Pd) and 7 (Pt) coded among the two lowest.

Because many of the sub-groups in this study (for both men and women) were very small, it was thought that Fisher's exact probability test might be helpful. Siegel[25] states, "The Fisher exact probability test is an extremely useful nonparametric technique for analyzing discrete data (either nominal or ordinal) when the two independent groups are small in size and when the scores from the two groups fall into one or the other of two mutually exclusive classes." The determining of the exact probabilities by means of Fisher's formula is rather laborious, but many of the values may be obtained from a "Table of Critical Values of D in the Fisher-Yates Test."[26] The Fisher test did not add any new information for this study; but it did support some of the previously determined relationships.

Summary

In summary, it can be stated that the analysis of the data for this study indicates the following relationships, —all of which are limited to the population studied or a similar group:

1. A high coding of scale 3 (Hy) for women teachers was positively related to a supervisory rating of outstanding. Although a statement of this relationship does not appear in any of the literature, Professor L. E. Drake, Head of the Student Counseling Center of the University of Wisconsin, has stated orally that in his experience he has found such a relationship to exist:

Thirty-five percent of the small group of men in the study had scale 3 (Hy) coded among the three highest scales. All of these men were rated above average or outstanding by supervisors.

2. Scale 7 (Pt), although coded highest or among the three highest scales for women teachers more frequently than any other scale, did not distinguish to a marked degree between good and poor teachers as indicated by supervisory ratings.

An absence of scale 7 coded high seemed to be somewhat related to high supervisory ratings for women teachers.

Low codings of scale 7 appeared to be positively related to low supervisory ratings for women.

Scale 7 was not coded high for men as frequently, relatively speaking, as for women teachers. It was coded among the lowest two scales more frequently for men than were many of the other scales.

3. For the women teachers in this study, a high coding of scale 5 (Mf) was positively related to supervisory ratings which were above average, whereas a low coding of this scale seemed to be somewhat related to average or below supervisory ratings. There are some indications that more extensive data might reveal more definite relations between high or low codings of scale 5 and supervisory ratings. It can be noted that in the lowest rated group of female teachers, there is no occurrence of scale 5 coded high.

For male teachers, scale 5 (Mf) coded highest appeared in 45 percent of the cases. It was coded among the three highest scales in 65 percent of the cases. Only one man had scale 5 coded among the two lowest scales.

4. A low coding of scale 2 (D) for women teachers was associated with supervisory ratings that were above average.

5. Scale 0 (Si) was coded lowest much more frequently than it was coded highest for women teachers. A low coding seemed to be somewhat related to above average supervisory ratings.

Scales 2 (D) and 5 (Mf) were also coded lowest more frequently than highest for women.

6. There were only four out of 147 female

teachers who had no scales coded low, whereas five of the twenty male teachers had no scale coded low.

It was pointed out in an earlier section that the MMPI was originally designed for and standardized on mental patients, but that it has since been used advantageously on many different more normal groups in a variety of situations.

It should be pointed out that the people in this study, who were given the MMPI, were all University of Wisconsin graduates who were granted a certificate to teach. The University of Wisconsin maintains an extensive testing, counseling, and guidance program and it is more than likely that most of the students with serious personality abnormalities who were not eliminated through the natural processes associated with attending college would be eliminated through the testing and counseling program. At the University of Wisconsin, students who deviate seriously from the normal are sent to the University psychiatrist for counseling or treatment.

Apparently the University of Wisconsin program is effective in eliminating the serious personality cases before they are graduated and sent out to teach. For the 167 cases in this study, there were only two reports from supervisors or principals which were concerned with emotional instability which was serious enough to affect the teacher's work. If the University's program is successful, one would not expect the MMPI to discover many extreme deviations from normality among the group in this study.

The MMPI may be a valuable instrument in helping to eliminate the occasional seriously deviated personality that does graduate and seek a teaching position.

Further experimentation with the MMPI is recommended to see if there isn't some "cut off" point on certain scales, possibly scales 0 (Si), 5 (Mf), and 7 (Pt), or on combinations of scales, a score beyond which would eliminate a student from teacher candidacy. In order to find such points, especially if combinations of scales are to be considered, it would probably be necessary to carry on a study of several years duration in order to get enough individuals in the study.

One serious deficiency or fault is frequently enough to make a person unsuccessful. Success may be due to one or more superiorities, providing a serious deficiency does not offset the superior qualities, or to a number of mediocrities.

The various scales of the MMPI have been referred to frequently by number or by number and abbreviation. It has been pointed out that the scales were first developed for clinical use with hospital patients and they bear titles indicative of deviate behavior. With the increased use of the Inventory for more normal people and further investigation in regard to the interpretation of scores on the various scales, the authors of the Inventory have suggested that the clinical titles and descriptions of the various scales should probably not be used. They suggest calling the scales by number or letter. This has been done throughout this report.

However, for those who are not familiar with the MMPI, brief descriptions of the various scales will be given. This will be done primarily in the language of Drake[27] who describes the scales in terms of the behavior of more normal persons.

Scale 0 (Si) has the title "Social Introversion-Extroversion". This scale was derived and cross-validated by Drake[28,29]. Its title is more or less self explanatory. The scale is intended to measure a characteristic that is thought to be important in college adjustment. Personality patterns with a high coding of scale 0 (Si) have been found among persons showing introvertive characteristics, especially shyness, social insecurity, and social withdrawal. One might expect such a personality characteristic to be detrimental to teaching success. On the other hand, a low coding of scale 0 (Si), which is probably indicative of an adequate social adjustment, may indicate a higher degree of teacher effectiveness. The present study found low codings of scale 0 (Si) for women teachers was associated with above average supervisory ratings.

Scale 1 (Hy) bears the name "Hypochondriasis". Elevation of this scale appeared relatively infrequently with college men and women, probably because persons with problems of a nature which caused a high coding of this scale sought medical aid. A high coding of this scale in a teacher's profile might indicate trouble which needs medical treatment.

The "Depression" scale has the number 2. A high score on scale 2 (D) is indicative of unhappiness and depression. This condition may be associated with problems in the area of human relations, lack of success in a chosen endeavor, or dissatisfaction with some aspect of the immediate environment. Scale 2 coded low with women were related to extroversion or socially outgoing characteristics. The study here reported found low codings of scale 2 were associated with above average supervisory ratings for women.

The operation of scale 3, "Hysteria" (Hy), with college students appeared to be different than with the patients upon whom it was derived. For normal college people, scale 3 coded high appears to be indicative of a lack of social problems. Scale 3 coded high along with certain other scales seems to indicate a variety of tendencies, but in most of them, a good general adjustment seems prevalent. This should contribute to a high degree of teacher effectiveness. The findings of the present study were that such a relationship exists.

The "Psychopathic Deviate" scale, number 4 (Pd), seemed to have little to do with college ed-

ucated people. Probably persons who are psychopathic deviates get into trouble and are eliminated. Elevation of this scale is associated with delinquency. There is some evidence that scale 4, in combination with other scales, often is associated with antagonism to authority, aggressive behavior, and conflict situations when it is elevated in a man's profile. With women this aggressiveness seems to be expressed in home conflict or in socially outgoing behavior depending upon the rest of the profile.

Scale 5 (Mf) was the least well validated of the basic scales on the MMPI. It bears the title "Masculinity-Femininity" and was designed to assist in the identification of homosexuals. However, the scale failed to distinguish between the homosexual group and normals. A high coding of this scale among men is particularly frequent in an educational setting. It is a complicated scale to interpret, but it is one of the most valuable scales in counseling students. An elevation of scale 5 means different things, depending upon what other scales are elevated with it. A low coding of scale 5 for women was very common. This corresponds to a high coding for men. In many cases a low coding of 5 acted as a suppressor variable for behavior which is suggested by the rest of the profile. In some patterns it appears to intensify or support the rest of the profile. A high coding of scale 5 with women appears to be associated with socially outgoing behavior. This should contribute to teacher effectiveness.

A high coding of scale 6, "Paranoid", seems to occur infrequently with college men except when paired with scale 4, "Psychopathic Deviate". In this case, it differentiated the anxiety sub-group. High coding of scale 6 (Pa) may be an indicator of personal sensitivity and perhaps extreme concern about the reactions of others to something perceived by the counselee as a deficiency. With teachers a high coding of scale 6 may indicate a person who is overly sensitive to the reactions of fellow teachers or even to students.

Scale 7 (Pt) bears the name "Psychasthenia". A high coding of scale 7 appears to be related to general anxiety states. A low coding with men probably indicates good adjustment. With women a high coding of scale 7 also suggests a feeling of insecurity. In general, high codings of this scale suggest personal problems, insecurity, state of anxiety, physical complaints—none of which are conducive to good teaching and personal relationships. The present study found an absence of scale 7 coded high was somewhat related to high supervisory ratings.

Scale 8, "Schizophrenia", when coded high, frequently suggests mental disorganization, vague goals, lacking motivation, general confused state. This appeared to be more serious with women than with men. It would seem that conditions associated with scale 8 (Sc) are a hindrance to effective teaching.

The "Hypomania" scale, number 9 (Ma), when coded high, appeared to indicate aggressiveness or belligerence in men and energetic action of a more socially approved nature in women. The high coding seemed to be modified considerably depending upon which scales were coded with it. For example, a high coding of scale 0 (Si) with a high coding of 9 was associated with social shyness, but when 0 (Si) was low with 9 high, socially outgoing and verbal behavior was suggested. Low codings of scale 9 (Ma) for women were usually indicative of social shyness when either scale 0 (Si) or scale 2 (D) was coded low. In general, the type of personality to be expected in a teacher with scale 9 coded either high or low seems to depend upon what scales are associated with it and in what manner they are related.

It should be emphasized that it is dangerous to associate the clinical names of most of these scales with individuals who may possess a high or low score on them. Perhaps new scales with new names should be developed for use when normal persons are being studied.

SECTION IV

Summary and Conclusions

Statement of the Problem

The purpose of this study was to investigate the relationship, if any, between scores and profile patterns on the MMPI and success in teaching as indicated by supervisory ratings. Personality has been considered by many to be an important factor in a teacher's success. The many aspects of personality are combined in a multitude of ways to make people react differently to various situations. It is thought that some assessment of the different aspects of a teacher's personality may be helpful in predicting that person's success or failure as a teacher. This study was an attempt to discover whether or not there were relationships between aspects of a teacher's personality and his professional success.

Subjects and Instruments

This study was concerned with students who were enrolled as freshmen at the University of Wisconsin during the fall semesters of 1949, 1950, and 1951, who received the University Teacher's Certificate in 1953, 1954, 1955, or 1956, and who subsequently taught in the public schools of Wisconsin. One hundred sixty-seven subjects constituted this population. Only twenty of these persons were males.

Supervisory ratings were obtained for these individuals by submitting a letter and a rating

blank to the superintendents, principals, and supervisors of the schools in which they taught.

Measures of the various phases of the teacher's personality were obtained from the results of the group form of the MMPI which had been given to the subjects when they were freshmen. The scores had been coded according to a plan recommended by one of the authors of the Inventory. The MMPI is a widely used instrument that has been found valuable in clinical studies. It is important that its value for this particular purpose should be ascertained.

Method Employed

Although several different statistical techniques were tried with the data of this study, the chi-square test seemed the most appropriate and provided the best analysis. For purposes of investigation the null hypothesis that teachers with different personality profiles would not differ in their supervisory ratings was established. The chi-square technique was used to determine if any particular sub-group differed from the whole population or from certain other sub-groups in respect to supervisory ratings. Since this was not a sampling study, no generalizations to a larger group could be made, but the findings could be useful in predicting the success of future classes of teachers if one were willing to consider the population of this study as a sample of the population of future teachers, or that conditions will remain basically constant.

Comparisons of frequencies, percentages and other aspects of the data which seemed pertinent were also made. Since the number of males in the study was small and it is generally conceded that males differ from females in their personality patterns, the analysis of the data for the male teachers was made separately. This was limited to the type of descriptive analysis mentioned at the beginning of this paragraph.

Findings

The findings of this study briefly summarized were:

1. There were differences in personality patterns among female teachers with different supervisory ratings. A high coding of scale 3 (Hy) for women was positively related to a supervisory rating of outstanding effectiveness. Scale 3 was also coded high very frequently for male teachers, but the number of males was too small to make any conclusions regarding the relationship.

2. Scale 5 (Mf) seems to be positively correlated to supervisory ratings for women teachers. The findings in this respect were not very significant statistically. Scale 5 coded high was the most frequently noted aspect of the male personality profiles. It was coded among the highest three scales 65 percent of the time.

3. There appears to be an inverse relationship between scale 2 (D) and supervisory ratings for women teachers.

4. Scales 0 (Si), 2 (D), and 5 (Mf) were coded lowest more frequently than they were coded highest for the women teachers.

5. Women teachers tended to have scales coded low more frequently than men teachers.

Conclusions

If one wishes to consider the group here studied as representative of future groups of prospective teachers, one might conclude that the MMPI has good possibilities for helping in the prediction of teacher success. While the study has been limited by small numbers in many of the sub-groups, nevertheless, differences in personality profiles did occur for the different supervisory categories.

Although relatively small samples are frequently adequate in statistical studies, in investigations of this nature, because of the large number of sub-groups into which the total group is broken down, the sub-groups are frequently too small to permit of statistical treatment. Such an investigation demands extremely large numbers if a statistical study is to be made. Such studies might be carried on over a period of years and with the cooperation of several large teacher-training institutions.

The natural processes of college life and the efforts of testing, counseling, and guidance personnel tend to eliminate many of the persons with extreme personality deviations from the rolls of candidates for teaching positions. The MMPI can be of help in identifying these individuals. It has potential usefulness in aiding in the prediction of the success of those who do accept teaching positions.

FOOTNOTES

1. Rostker, L. E. "The Measurement of Teaching Ability," Journal of Experimental Education, XIV (1945), p. 50.
2. Lins, L. J. "The Prediction of Teaching Efficiency," Journal of Experimental Education, XV (1946), p. 59.
3. Lins, op. cit.
4. LaDuke, C. V. "The Measurement of Teaching Ability," Journal of Experimental Education, XIV (1945), p. 100.
5. Albert Ellis, Consulting Psychologist, 333 W. 56th Street, New York 19, New York.
6. Buros, Oscar F., (Editor), Fifth Mental Measurements Yearbook (Highland Park, N.J.: The Gryphon Press, 1959), p. 166.
7. Drake, L. E. and Oetting, E. R. An MMPI

Codebook for Counselors (Minneapolis: University of Minnesota Press, 1959), p. 10.
8. Hathaway, J. R., and Monachesi, E. D. Analyzing and Predicting Juvenile Delinquency with the MMPI (Minneapolis: University of Minnesota Press, 1953), pp. 13-28.
9. Lins, op. cit.
10. Hathaway, S. R. "A Coding System for MMPI Profiles," Journal of Consulting Psychology, XI (1947), pp. 334-37.
11. Hathaway, op. cit.
12. Du Mas, F. M. "On the Interpretation of Personality Profiles," Journal of Clinical Psychology (1947), pp. 57-65.
13. McCall, William A., and Krause, Gertrude R., "Measurement of Teacher Merit for Salary Purposes," Journal of Educational Research, LIII (1959), pp. 73-5.
14. Tate, Merle W., and Clelland, Richard C., Nonparametric and Shortcut Statistics in the Social, Biological, and Medical Sciences (Danville, Ill.: Interstate Printers and Publishers, 1957), p. 52.
15. Walker, Helen and Lev, Joseph. Statistical Inference (New York: Holt, 1953), p. 107.
16. Tate and Clelland, op. cit., p. 71.
17. Cochran, W. G., "The Comparison of Percentages in Matched Samples," Biometrika, XXXVII (1950), pp. 256-66.
18. Lindquist, E. F. Statistical Analysis in Educational Research (Chicago: Houghton Mifflin Co., 1940), pp. 44-5.
19. An explanation of this method of calculating chi-square may be found in the Appendix.
20. Walker and Lev, op. cit., pp. 105-6.
21. Drake and Oetting, op. cit., pp. 24-5.
22. Ibid., p. 19.
23. Ibid., pp. 22-3.
24. Ibid., p. 15.
25. Siegel, Sidney, Nonparametric Statistics for the Behavioral Sciences (New York: McGraw Hill, 1956), pp. 96 ff.
26. Finney, D. J. "The Fisher-Yates Test of Significance in 2 × 2 Contingency Tables," Biometrika, XXXV (1948), pp. 149-54.
27. Drake and Oetting, op. cit.
28. Drake, L. E., "Scale 0 (Social Introversion)," in Basic Readings on the MMPI in Psychology and Medicine, G. S. Welsh and W. G. Dahlstrom, editors (Minneapolis: University of Minnesota Press, 1956).
29. Drake, L. E., and Thiede, W. B. "Further Validation of Scale 0 (Si)," in Basic Readings on the MMPI in Psychology and Medicine, G. S. Welsh and W. G. Dahlstrom, editors (Minneapolis: University of Minnesota Press, 1956).

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 4, June 1961)

A COMPARISON OF TEACHERS' RESPONSE PATTERNS ON THE MMPI WITH RESPONSE PATTERNS OF SELECTED NON-TEACHER GROUPS

ELMER CHRISTIAN RIECK*
Wisconsin State College
Eau Claire, Wisconsin

SECTION I

The Problem and Its Importance

Statement of the Problem

THE PURPOSE of this study was to identify and then compare profile patterns of teachers on the group form of the Minnesota Multiphasic Personality Inventory, hereafter referred to as the MMPI, with the profile patterns of selected non-teacher groups of students at the University of Wisconsin. The three main groups studied, divided according to sex, were students who were graduated from the University of Wisconsin with teaching certificates, students with majors in Letters and Science who did not prepare for teaching, and freshmen and sophomores who dropped out prior to starting the junior year. Those who prepared for teaching were further subdivided into academic and special field majors.

Specifically, the researcher will deal with the following four questions:

1. What, if any, MMPI profile patterns are found for students who were graduated from the University of Wisconsin with teaching certificates?
   a. Graduates of the School of Education with academic majors.
   b. Graduates with teaching certificates in special fields—Home Economics, Music, Agriculture, Art Education, and Physical Education.
2. What, if any, MMPI profile patterns are found for students who were graduated from the University of Wisconsin from the College of Letters and Science but did not prepare for teaching?
3. What, if any, MMPI profile patterns are found for students who dropped out of the University prior to starting the junior year?
4. If profile patterns are found, how do they compare among these groups?

It is hoped that the results of the study may prove useful in counseling students in deciding between teaching and other non-teaching careers.

The writer became interested in the MMPI when in student personnel work during the period of 1948-58. During these years he was privileged to see and counsel many students who sought guidance and who requested assistance in personnel, academic, social, and vocational problems. It was his primary function to help facilitate the growth and development of the student of a realistic understanding of himself and his relationship with society.

In the discharge of such functions the use of psychometric data may be very helpful. On the other hand, they may lead, at times, to erroneous conclusions. The fact that psychometric data are often useful in many situations must grow out of their basic validity for the purpose used.

While most of the earlier researches on the MMPI were done in a hospital setting, research has been conducted in a variety of situations with more normal persons. Drake and his associates (10) have made a number of studies with university

*Only through the cooperation of many people was this study possible. My sincere gratitude and appreciation is expressed to Professor A. S. Barr, the writer's major professor and Chairman of the Reading Committee, and Professors L. E. Drake and H. J. Klausmeier, members of the Reading Committee. They have been extremely generous in giving their time and effort to this investigation. Their continuous guidance, constructive criticism and encouragement is deeply appreciated. My sincere gratitude is also extended to my friends and associates who have encouraged and aided in this study.

students.

Importance of the Problem

At the present time nearly every large university and most of the smaller institutions offer their students some type of guidance service. At the University of Wisconsin, for example, the Student Counseling Center has offered this service to the students since the Center opened in 1944.

One of the important functions of these counseling centers is that of academic and vocational guidance. The work of assisting students in selecting their academic and vocational goals is difficult. Because of the demands on the part of the students for this type of help and the desire on the part of the counselors to be effective, it was necessary that the MMPI be carefully studied so that it may be used as effectively as possible.

The present study was designed to obtain more information about the suitability of using the MMPI with University of Wisconsin students in academic and vocational guidance, especially in the selection of future teachers. The MMPI is a widely used instrument and has been found useful in clinical studies. It is important that its value for this particular purpose should be ascertained.

SECTION II

The Procedures Used in the Investigation

The Design of the Study

The study here reported was of the comparative-descriptive type in which groups of prospective teachers and non-teaching groups were compared upon scales and profile patterns of the MMPI. In an exploratory study such as this, it seemed best to work with a well-defined population about whom precise and complete information could be had. The study concerned students enrolled at the University of Wisconsin during the fall semesters of 1949, 1950, and 1951.

The Minnesota Multiphasic Personality Inventory

The MMPI was chosen as the measuring instrument to be used in this study. It is normally administered to students entering as freshmen during freshmen orientation week at the University of Wisconsin and the profile codes are entered on the cards by the staff of the Student Counseling Center. Profiles and other data were thus obtained for the subjects upon the basis of which comparisons were made of the several profile patterns.

The MMPI is a psychometric instrument developed at the University of Minnesota. It is a product of S. R. Hathaway, Professor and Director of the Division of Clinical Psychology, University of Minnesota, and J. C. McKinley (deceased) formerly Head of the Department of Neuropsychiatry, University of Minnesota. The Inventory was designed to provide, in a single test, scores on all the more important clinical phases of personality. It seemed that a multiphasic personality inventory would be best suited for our purposes.

The instrument consists of 550 different statements covering a wide range of subject matter, from the physical condition of the individual being tested to his morale and attitude (15). In administering the test, the subject is asked to respond to all statements, which are in the first person, as True, False, or Cannot Say. After he has responded to all items in this manner, the responses are scored and tabulated to faciliate this part of the operation. Drake, at the University of Wisconsin Counseling Center, has recorded the significant responses on an IBM machine score sheet. This device saves considerable time and labor in scoring. Experimentation has shown that it is adapted to useful classification and understanding of college students.

Subjects and Data-Gathering Procedures

The students for this study were students at the University of Wisconsin who entered as freshmen in the fall semesters of 1949, 1950, and 1951.

A list of the students who were graduated in 1953, 1954, and 1955 with teaching certificates was obtained from the Office of the Dean of the School of Education. The profiles of all these students who had taken the MMPI in 1949, 1950, and 1951 were used in this study. Of the total of 376 subjects, 99 were males and 277 females. The Korean situation and the fact that a larger percentage of females than males go into the teaching profession accounts for the larger female group.

The total population consisted of 1407 students, 654 males and 753 females. The profile patterns and other data were obtained from the files of the Student Counseling Center.

Graduates from the College of Letters and Science and freshman and sophomore drop-outs were secured from the records on file at the Alumni Records Office. The profiles of all the Letters and Science students who had taken the MMPI in 1949, 1950, and 1951 that were readily available, 278 males and 151 females, compromises the population of this group of approximately 1750. A total of 502 subjects, 277 males and 225 females, were used from the freshman and sophomore drop-out group. This represented all subjects for whom both drop-out records and MMPI's were available.

Inasmuch as this study attempted to identify and then compare profile patterns of students who subsequently become teachers with profile patterns of other selected groups of students, the subjects were divided into four groups as follows:

1. University of Wisconsin graduates with teaching certificates in the academic subject fields.
2. Graduates with teaching certificates in special fields—Home Economics, Music, Agriculture, Art Education, and Physical Education.
3. Students who were graduated from the University of Wisconsin from the College of Letters and Science but did not prepare for teaching.
4. Students who dropped out of the University prior to starting the junior year.

In order to identify common profile patterns among individuals within a group and to compare possible profile patterns among groups, each group must show relatively homogenous behaviors on the MMPI. Therefore, the group must be selected on the basis of as many common underlying traits as possible. In this study the groups were divided in accordance to sex and education and/or vocational goals.

Coding of the Data

The coding system of the profile analysis was that reported by Hathaway (14). The scales were assigned numbers as follows:

0	Si	- Social Introversion
1	Ha	- Hypochondriasis
2	D	- Depression
3	Hy	- Hysteria
4	Pd	- Psychopathic Deviate
5	Mf	- Masculinity-Femininity
6	Pa	- Paranoid
7	Pt	- Psychasthenia
8	Sc	- Schizophrenia
9	Ma	- Hyponrenia
X	None	

All the scales with T scores above 54 were coded high and arranged according to magnitude. Only the three highest scales were used in this study. The scales with T scores of 45 or less were coded low with a maximum, in this study, of two being coded low except in cases where there were ties in which case the first three low scales were included.

Tabulating the Data

The frequency of occurrence of each of the following patterns of scales were tabulated (8):

1. The frequency of each scale coded high among the three highest scales.
2. The frequency of each scale coded low among the two lowest scales. (In case of a tie, the first three low scales were included.)
3. The frequency of each pair of scales coded among the three highest scales.
4. The frequency of each scale coded high of the three highest with each scale coded low of the two lowest.
5. The frequency of each pair of high scales coded among the three highest with each scale coded low of the two lowest.

Comparisons of the profile patterns among the main groups of students were made by Yule's formula for differentiation of groups (25). This formula assumes randomly drawn groups. Strictly speaking, these groups were not randomly drawn. However, for some larger group of Wisconsin graduates, the groups might be regarded as random groups. To make such a statement it must be assumed that conditions remain unchanged in all essential conditions and that these conditions persist for some period of time; also that the types of individuals appearing in the data have an equal opportunity to appear at each successive drawing. The data here are for graduates who entered as freshmen in three separate years, namely, 1949, 1950, and 1951. The statistical test employed in the study answers the question: What are the probabilities that differences as great or greater than those found might have arisen by chance?

SECTION III

Statistical Analysis and Findings

Introduction

The purpose of this study, as stated in Section I, was to identify and compare profile scales and patterns of teachers on the group form of the MMPI with the profile patterns of selected non-teacher groups of students at the University of Wisconsin. The following four questions were asked:

1. What, if any, MMPI profile patterns are found for students who graduated from the University of Wisconsin with teaching certificates?
   a. Graduates of the School of Education with academic majors.
   b. Graduates with teaching certificates in special fields—Economics, Music, Agriculture Art Education, and Physical Education.
2. What, if any, MMPI profile patterns are found for students who were graduated from the col-

lege of Letters and Science but did not prepare for teaching?
3. What, if any, profile patterns are found for students who dropped out of the University prior to starting the junior year?
4. If profile patterns are found, how do they compare among these groups?

In order to answer these questions it is necessary to explain how the data were analyzed and what the statistical findings were.

Yule's Formula for Differentiations of Groups

In Section II, The Design of the Study, the MMPI, the coding of the data, the procedure for tabulating the data, were explained. These data were analyzed through the use of Yule's (25) formula for differentiation of groups. The significant difference between the groups was determined by the use of the formula for the standard error of differences.

$$e_{12}2 = \frac{p_1 q_1}{n^1} + \frac{p_2 q_2}{n^2} \qquad C.R. = \frac{d}{e}$$

where

e_{12} is the standard error of difference

p_1 = the percent of our group having the given code type
q_1 = the percent of the group not having the given code type
p_2 = the percent of the other group having the given code type
q_2 = the percent of that group not having the given code type
n_1 and n_2 = the number of individuals in each group (25)

The size of $e_{12}2$ is an index of the probability of the difference in proportions arising through chance due to the fact that the researcher sampled the total number of persons engaged as explained in Section II. According to Yule, if the difference in proportion does not exceed three times the error of the difference, it may be obliterated by an error of simple sampling on taking fresh samples in the same way from the same material. But for practical purposes the probability of the difference being obliterated in this manner is very little less if the error is twice the difference in proportion.

Six group comparisons were made for each sex, male and female, as follows:

1. Graduates with teaching certificates in the academic subject fields were compared with the graduates with teaching certificates in special fields.
2. Graduates with teaching certificates in the academic subject fields were compared with students who were graduated from the College of Letters and Science but did not prepare for teaching.
3. Graduates with teaching certificates in the academic subject fields were compared with students who dropped out of the University prior to starting the junior year.
4. Graduates with teaching certificates in special fields were compared with students who were graduated from the College of Letters and Science but did not prepare for teaching.
5. Graduates with teaching certificates in special fields were compared with the students who dropped out of the University prior to starting the junior year.
6. Students who were graduated from the College of Letters and Science but did not prepare for teaching were compared with the students who dropped out of the University prior to starting the junior year.

In this study a difference in percent was considered significant by the researcher if it was two or more times as great as its standard error, i.e., it is significant beyond the 5 percent level.

Statistical Findings

The statistical findings are summarized in the tables that follow. They include results for those scales and patterns for which the difference between the groups compared is two or more times as great as the standard error (9). Code types refer to the coding of the scales, either high or low, and combinations of scales, paired high and high with low.

In this investigation, combination of scales paired high with low scales were not tested for significance. The frequencies of occurrence for this code type appear in the Appendix (which will be found in original thesis on file in Library of University of Wisconsin). It was felt that there was an insufficient number of subjects to allow tests for significant differences between these very low frequencies. For the interested researcher some statistical comparisons could be made and other information gained from inspection, if so desired. Drake (9) has reported the frequency of occurrence of a number of patterns found in male clients.

In reporting the findings, the tables show the percentage of the frequency of occurrence of the various code types for all sub-groups compared with the corresponding percentage of the frequency of other sub-groups along with the critical

TABLE I

COMPARISON OF GRADUATES WITH TEACHING CERTIFICATES IN THE ACADEMIC SUBJECT FIELDS WITH THE GRADUATES WITH TEACHING CERTIFICATES IN SPECIAL FIELDS—MALES

Code Types	Percent of Academic Group	Percent of Special Fields Group	Critical Ratio
X High	14. 3*	00. 0	2. 42
7 Low	08. 6*	26. 6*	2. 47
13 Paired High	00. 0	07. 8	2. 33
46 Paired High	00. 0	09. 4*	2. 58
2-9 High with Low	00. 0	06. 3	2. 07
3-2 High with Low	00. 0	07. 8*	2. 33
4-8 High with Low	00. 0	10. 9*	2. 79
6-7 High with Low	00. 0	06. 3	2. 07
Number in each group	35	64	

*These items represent samples of 5 or more cases in one or the other of the compared group N. B. after the number in each case.

TABLE II

COMPARISON OF GRADUATES WITH TEACHING CERTIFICATES IN THE ACADEMIC SUBJECT FIELDS WITH THE GRADUATES WITH TEACHING CERTIFICATES IN SPECIAL FIELDS—FEMALES

Code Types	Percent of Academic Group	Percent of Special Fields Group	Critical Ratio
0 Low	27. 4*	16. 8*	2. 14
5 Low	42. 1*	30. 1*	2. 07
12 Paired High	03. 0*	00. 0	2. 26
0-3 High with Low	03. 0*	10. 6*	2. 38
4-6 High with Low	03. 0*	00. 0	2. 26
5-1 High with Low	02. 4	09. 7*	2. 41
9-5 High with Low	16. 5*	08. 0*	2. 20
Number in each group	164	113	

*These items represent samples of 5 or more cases in one or the other of the compared group N. B. after the number in each case.

TABLE III

COMPARISON OF GRADUATES WITH TEACHING CERTIFICATES IN THE ACADEMIC SUBJECT FIELDS WITH STUDENTS WHO WERE GRADUATED FROM THE COLLEGE OF LETTERS AND SCIENCE BUT DID NOT PREPARE FOR TEACHING—MALES

Code Types	Percent of Academic Group	Percent of Letters and Science Group	Critical Ratio
X High	14. 3*	0. 7	2. 28
8 High	42. 9*	25. 2*	2. 02
0 Low	31. 4*	58. 9*	3. 28
05 Paired High	00. 0	03. 6*	3. 21
13 Paired High	00. 0	07. 2*	4. 65
14 Paired High	00. 0	03. 6*	3. 21
36 Paired High	02. 7	11. 2*	2. 43
46 Paired High	00. 0	03. 6*	3. 21
56 Paired High	02. 9	11. 5*	2. 51
67 Paired High	00. 0	02. 2*	2. 49
0-X High with Low	00. 0	01. 8*	2. 25
1-0 High with Low	00. 0	06. 5*	4. 39
1-7 High with Low	00. 0	02. 5*	2. 66
1-9 High with Low	00. 0	02. 5*	2. 66
2-9 High with Low	00. 0	03. 6*	3. 21
3-1 High with Low	00. 0	05. 8*	4. 14
3-2 High with Low	00. 0	05. 0*	3. 82
3-4 High with Low	00. 0	02. 5*	2. 66
4-0 High with Low	05. 7	16. 9*	2. 47
4-8 High with Low	00. 0	01. 8*	2. 25
4-9 High with Low	00. 0	07. 7*	4. 81
6-0 High with Low	02. 9	10. 1*	2. 14
6-7 High with Low	00. 0	05. 4*	4. 00
7-0 High with Low	00. 0	04. 0*	3. 39
7-1 High with Low	00. 0	02. 5*	2. 66
7-9 High with Low	00. 0	04. 0*	3. 39
8-1 High with Low	00. 0	01. 8*	2. 25
8-6 High with Low	00. 0	03. 6*	3. 21
9-3 High with Low	00. 0	02. 2*	2. 49
9-4 High with Low	00. 0	03. 6*	3. 21
Number in each group	35	278	

*These items represent samples of 5 or more cases in one or the other of the compared group N. B. after the number in each case.

TABLE IV

COMPARISON OF GRADUATES WITH TEACHING CERTIFICATES IN THE ACADEMIC SUBJECT FIELDS WITH STUDENTS WHO WERE GRADUATED FROM THE COLLEGE OF LETTERS AND SCIENCE BUT DID NOT PREPARE FOR TEACHING—FEMALES

Code Types	Percent of Academic Group	Percent of Letters and Science Group	Critical Ratio
3 High	28.0*	39.1*	2.09
5 Low	42.1*	26.5*	2.96
78 Paired High	20.1*	11.3*	2.17
1-6 High with Low	02.4	00.0	2.00
2-9 High with Low	04.9*	00.0	2.90
3-X High with Low	02.4	06.6*	2.03
3-1 High with Low	02.4	07.9*	2.20
4-5 High with Low	13.4*	06.0*	2.25
4-9 High with Low	04.9*	0.7	2.32
7-X High with Low	04.9*	12.6*	2.42
7-0 High with Low	06.7*	02.0	2.08
7-5 High with Low	22.0*	12.6*	2.24
8-5 High with Low	15.9*	07.9*	2.22
8-9 High with Low	04.9*	0.7	2.31
Number in each group	164	151	

*These items represent samples of 5 or more cases in one or the other of the compared group N. B. after the number in each case.

TABLE V

COMPARISON OF GRADUATES WITH TEACHING CERTIFICATES IN THE ACADEMIC SUBJECT FIELDS WITH STUDENTS WHO DROPPED OUT OF THE UNIVERSITY PRIOR TO STARTING THE JUNIOR YEAR—MALES

Code Types	Percent of Advanced Group	Percent of Drop-Outs	Critical Ratio
X High	14. 3*	02. 2*	2. 02
7 Low	08. 6	23. 5*	2. 77
02 Paired High	00. 0	02. 9*	2. 87
05 Paired High	00. 0	01. 8*	2. 25
09 Paired High	00. 0	01. 8*	2. 25
13 Paired High	00. 0	04. 7*	3. 69
14 Paired High	00. 0	01. 8*	2. 25
34 Paired High	02. 9	09. 7*	2. 03
37 Paired High	00. 0	02. 2*	2. 49
38 Paired High	00. 0	02. 5*	2. 66
46 Paired High	00. 0	05. 1*	3. 83
49 Paired High	05. 7	15. 2*	2. 24
67 Paired High	00. 0	04. 0*	3. 39
9X Paired High	00. 0	01. 8*	2. 25
0-X High with Low	00. 0	01. 8*	2. 25
0-1 High with Low	00. 0	02. 2*	2. 49
0-2 High with Low	00. 0	02. 2*	2. 49
0-5 High with Low	00. 0	02. 2*	2. 49
0-6 High with Low	00. 0	03. 6*	3. 21
1-2 High with Low	00. 0	02. 5*	2. 66
1-9 High with Low	00. 0	02. 2*	2. 49
2-9 High with Low	00. 0	04. 3*	3. 52
3-2 High with Low	00. 0	04. 3*	3. 52
3-5 High with Low	00. 0	03. 2*	3. 02
4-5 High with Low	00. 0	02. 2*	2. 49
4-8 High with Low	00. 0	05. 1*	3. 83
4-9 High with Low	00. 0	02. 9*	2. 87
5-3 High with Low	00. 0	03. 2*	3. 02
6-7 High with Low	00. 0	05. 1*	3. 83
6-9 High with Low	00. 0	03. 2*	3. 02
7-1 High with Low	00. 0	02. 9*	2. 87
7-5 High with Low	00. 0	02. 5*	2. 66
7-9 High with Low	00. 0	03. 6*	3. 21
8-X High with Low	28. 6*	12. 3*	2. 06
8-5 High with Low	00. 0	02. 2*	2. 49
8-6 High with Low	00. 0	03. 6*	3. 21
9-3 High with Low	00. 0	03. 6*	3. 21
9-4 High with Low	00. 0	01. 8*	2. 25
9-7 High with Low	02. 9	11. 9*	2. 62
Number in each group	35	277	

*These items represent samples of 5 or more cases in one or the other of the compared group N. B. after the number in each case.

TABLE VI

COMPARISON OF GRADUATES WITH TEACHING CERTIFICATES IN THE ACADEMIC SUBJECT FIELDS WITH STUDENTS WHO DROPPED OUT OF THE UNIVERSITY PRIOR TO STARTING THE JUNIOR YEAR—FEMALES

Code Types	Percent of Academic Group	Percent of Drop-Outs	Critical Ratio
5 Low	42.1*	26.7*	3.17
7 High	42.1*	32.0*	2.04
9 High	39.0*	50.7*	2.30
29 Paired High	00.0	02.2*	2.24
49 Paired High	07.3*	14.2*	2.23
59 Paired High	04.9*	11.1*	2.30
78 Paired High	20.1*	12.4*	2.00
4-1 High with Low	0.6	05.8*	3.11
6-0 High with Low	06.1*	12.9*	2.34
7-5 High with Low	22.0*	13.8*	2.09
7-9 High with Low	06.7*	01.3	2.57
8-5 High with Low	15.9*	08.9*	2.04
9-1 High with Low	04.3*	11.6*	2.74
9-7 High with Low	05.5*	11.6*	2.19
9-8 High with Low	00.0	03.6*	2.90
Number in each group	164	225	

*These items represent samples of 5 or more cases in one or the other of the compared group N. B. after the number in each case.

TABLE VII

COMPARISON OF GRADUATES WITH TEACHING CERTIFICATES IN SPECIAL FIELDS WITH STUDENTS WHO WERE GRADUATED FROM THE COLLEGE OF LETTERS AND SCIENCE BUT DID NOT PREPARE FOR TEACHING—MALES

Code Types	Percent of Special Fields Group	Percent of Letters and Science Group	Critical Ratio
0 Low	29. 7*	58. 9*	4. 52
8 Low	26. 6*	12. 6*	2. 38
08 Paired High	00. 0	03. 2*	3. 02
14 Paired High	00. 0	03. 6*	3. 21
19 Paired High	01. 6	07. 2*	2. 55
29 Paired High	00. 0	01. 8*	2. 25
39 Paired High	04. 7	13. 7*	2. 69
45 Paired High	04. 7	12. 9*	2. 47
0-X High with Low	00. 0	01. 8*	2. 25
1-0 High with Low	01. 6	06. 5*	2. 28
3-X High with Low	01. 6	05. 8*	2. 00
3-4 High with Low	00. 0	02. 5*	2. 66
3-6 High with Low	00. 0	02. 2*	2. 49
4-0 High with Low	03. 1	16. 9*	4. 42
4-6 High with Low	00. 0	03. 6*	3. 21
4-8 High with Low	10. 9*	01. 8*	2. 29
7-1 High with Low	00. 0	02. 5*	2. 66
8-3 High with Low	00. 0	02. 5*	2. 66
8-9 High with Low	00. 0	02. 5*	2. 66
9-X High with Low	04. 7	11. 2*	2. 00
9-0 High with Low	09. 4*	22. 7*	3. 00
9-3 High with Low	00. 0	02. 2*	2. 49
Number in each group	64	278	

*These items represent samples of 5 or more cases in one or the other of the compared group N. B. after the number in each case.

TABLE VIII

COMPARISON OF GRADUATES WITH TEACHING CERTIFICATES IN SPECIAL FIELDS WITH STUDENTS WHO WERE GRADUATED FROM THE COLLEGE OF LETTERS AND SCIENCE BUT DID NOT PREPARE FOR TEACHING—FEMALES

Code Types	Percent of Special Fields Group	Percent of Letters and Science Group	Critical Ratio
3 High	26.5*	39.1*	2.19
9 High	32.7*	49.7*	2.83
3 Low	13.3*	04.6*	2.40
04 Paired High	03.5	00.0	2.02
39 Paired High	05.3*	13.9*	2.44
5X Paired High	07.1*	01.3	2.24
67 Paired High	11.5*	04.6*	2.00
79 Paired High	10.6*	19.2*	2.20
89 Paired High	00.7	14.6*	3.50
4-6 High with Low	00.0	02.6	2.00
5-1 High with Low	09.7*	03.3*	2.06
5-3 High with Low	04.4*	00.0	2.28
6-0 High with Low	03.5	11.3*	2.52
7-3 High with Low	08.0*	01.3	2.47
7-5 High with Low	23.9*	12.6*	2.33
9-0 High with Low	08.8*	16.6*	2.25
9-7 High with Low	03.5	08.6*	2.17
Number in group	113	151	

*These items represent samples of 5 or more cases in one or the other of the compared group N. B. after the number in each case.

TABLE IX

COMPARISON OF GRADUATES WITH TEACHING CERTIFICATES IN SPECIAL FIELDS WITH THE STUDENTS WHO DROPPED OUT OF THE UNIVERSITY PRIOR TO STARTING THE JUNIOR YEAR—MALES

Code Types	Percent of Special Fields Group	Percent of Drop-Outs	Critical Ratio
X High	00.0	02.2*	2.49
3 Low	01.6	08.7*	3.07
8 Low	26.6*	13.7*	2.19
02 Paired High	00.0	02.9*	2.87
08 Paired High	00.0	03.6*	3.21
14 Paired High	00.0	01.8*	2.25
29 Paired High	00.0	03.0*	3.36
38 Paired High	00.0	02.5*	2.66
39 Paired High	04.7	11.9*	2.19
5X Paired High	00.0	01.8*	2.25
58 Paired High	15.6*	04.3*	2.40
0-X High with Low	00.0	01.8*	2.25
0-1 High with Low	00.0	02.2*	2.49
0-5 High with Low	00.0	02.2*	2.49
1-5 High with Low	00.0	01.8*	2.25
3-6 High with Low	00.0	03.6*	3.21
4-0 High with Low	03.1	11.6*	2.93
4-6 High with Low	00.0	02.9*	2.87
5-3 High with Low	00.0	03.2*	3.02
7-1 High with Low	00.0	02.9*	2.87
8-3 High with Low	00.0	04.0*	3.36
8-5 High with Low	00.0	02.2*	2.49
9-X High with Low	04.7	11.6*	2.11
9-3 High with Low	00.0	03.6*	3.21
9-6 High with Low	04.7	11.6*	2.11
Number in group	64	277	

*These items represent samples of 5 or more cases in one or the other of the compared group N. B. after the number in each case.

TABLE X

COMPARISON OF GRADUATES WITH TEACHING CERTIFICATES IN SPECIAL FIELDS WITH THE STUDENTS WHO DROPPED OUT OF THE UNIVERSITY PRIOR TO STARTING THE JUNIOR YEAR—FEMALES

Code Types	Percent of Special Field Groups	Percent of Drop-Outs	Critical Ratio
9 High	32. 7*	50. 7*	3. 25
0 Low	16. 8*	30. 2	2. 87
9 Low	19. 5*	09. 3*	2. 42
12 Paired High	00. 0	01. 8	2. 02
15 Paired High	00. 0	02. 2*	2. 24
16 Paired High	00. 0	02. 7*	2. 50
23 Paired High	00. 9	04. 9*	2. 35
3X Paired High	05. 3*	00. 9	2. 00
39 Paired High	05. 3*	11. 6*	2. 10
6X Paired High	00. 0	02. 7*	2. 03
64 Paired High	02. 7	07. 6*	2. 10
69 Paired High	00. 2	12. 9*	3. 98
0-9 High with Low	10. 0*	03. 1*	2. 40
1-7 High with Low	00. 0	01. 8	2. 02
4-1 High with Low	01. 8	05. 8*	2. 00
4-6 High with Low	00. 0	03. 1*	2. 70
5-9 High with Low	07. 1*	00. 9	2. 50
6-0 High with Low	03. 5	12. 9*	3. 33
6-1 High with Low	01. 8	06. 7*	2. 36
7-5 High with Low	23. 9*	13. 8*	2. 18
9-0 High with Low	08. 8*	20. 0*	2. 90
9-7 High with Low	03. 5	11. 6*	2. 96
Number in group	113	225	

*These items represent samples of 5 or more cases in one or the other of the compared group N. B. after the number in each case.

TABLE XI

COMPARISON OF STUDENTS WHO WERE GRADUATED FROM THE COLLEGE OF LETTERS AND SCIENCE BUT DID NOT PREPARE FOR TEACHING WITH THE STUDENTS WHO DROPPED OUT OF THE UNIVERSITY PRIO TO STARTING THE JUNIOR YEAR—MALES

Code Types	Percent of Letters and Science Group	Percent of Drop-Outs	Critical Ratio
1 High	18.9*	11.6*	2.13
5 High	54.7*	35.7*	4.57
0 Low	58.9*	23.8*	8.97
3 Low	03.6*	08.7*	2.51
5 Low	02.2*	07.2*	2.80
15 Paired High	05.8*	01.8*	2.50
19 Paired High	07.2*	03.2*	2.13
57 Paired High	14.0*	05.4*	3.47
58 Paired High	09.4*	04.3*	2.39
0-5 High with Low	00.0	02.2*	2.49
1-0 High with Low	06.5*	01.4	3.12
1-5 High with Low	00.0	01.8*	2.25
3-1 High with Low	05.8*	01.4	2.80
3-4 High with Low	02.5*	00.4	2.08
3-5 High with Low	00.4	03.2*	2.39
4-8 High with Low	01.8*	05.1*	2.13
4-9 High with Low	07.7*	02.9*	2.54
5-3 High with Low	00.4	03.2*	2.50
5-4 High with Low	04.0*	01.1	2.18
7-0 High with Low	04.0*	00.7	2.58
7-2 High with Low	00.4	02.9*	2.31
7-5 High with Low	00.0	02.5*	2.66
8-2 High with Low	02.2*	06.1*	2.31
9-0 High with Low	22.7*	15.5*	2.11
Number in group	278	277	

*These items represent samples of 5 or more cases in one or the other of the compared group N. B. after the number in each case.

TABLE XII

COMPARIOSN OF STUDENTS WHO WERE GRADUATED FROM THE COLLEGE OF LETTERS AND SCIENCE BUT DID NOT PREPARE FOR TEACHING WITH THE STUDENTS WHO DROPPED OUT OF THE UNIVERSITY PRIOR TO STARTING THE JUNIOR YEAR—FEMALES

Code Types	Percent of Letters and Science Group	Percent of Drop-Outs	Critical Ratio
3 High	39. 1*	27. 1*	2. 42
3 Low	04. 6*	10. 7*	2. 28
04 Paired High	00. 0	02. 7*	2. 50
15 Paired High	00. 0	02. 2*	2. 24
26 Paired High	00. 7	03. 6*	2. 07
2-9 High with Low	00. 0	03. 1*	2. 67
3-2 High with Low	10. 6*	03. 6*	2. 50
5-3 High with Low	00. 0	02. 2*	2. 24
5-9 High with Low	04. 6*	00. 9	2. 03
6-3 High with Low	00. 7	04. 0*	2. 24
7-X High with Low	12. 6*	05. 8*	2. 18
7-3 High with Low	01. 3	06. 2*	2. 63
7-5 High with Low	02. 6	13. 8*	3. 15
Number in group	151	225	

*These items represent samples of 5 or more cases in one or the other of the compared group N. B. after the number in each case.

ratio of the differences.

The data relative to male graduates with teaching certificates in the academic subject fields as compared with the graduates with teaching certificates in specific fields are given in Table I. Table II makes the same comparisons for females. Tables III and IV show the graduates with teaching certificates in the academic subjects as compared with the students who were graduated from the College of Letters and Science but did not prepare for teaching. Tables V and VI show the data for graduates with teaching certificates in the academic subject fields as compared with the students who dropped out of the University prior to starting the junior year.

Comparisons of graduates with teaching certificates in special fields with students who were graduated from the College of Letters and Science but did not prepare for teaching are shown in Tables VII and VIII. The same comparisons were made of the graduates in special fields with students who dropped out of the University prior to starting the junior year in Tables IX and X.

Tables XI and XII show the data in the same way for the students who were graduated from the College of Letters and Science but did not prepare for teaching as compared with the students who dropped out of the University prior to starting the junior year.

Perhaps the most important and highly significant finding is the fact that statistically significant differences did occur, i. e., personality differences as represented by significant differences in pattern of scales of the MMPI did occur in all groups.

At this stage of the research it is not possible to actually identify the meanings of these personality differences. Further research may identify the meaning of these code types in personality terms. This may point to a better idea of personality variables in the future.

It appears from the data that there were fewer differences among the graduates with teaching certificates in the academic fields as compared with the graduates with teaching certificates in the special fields than between any other groups. The two groups of graduates with teaching certificates seem to be relatively homogeneous. The fact that a grade point average of 2. 5 is required for admittance to the School of Education may possibly account for this finding together with the possibility that the personality characteristics that helped lead to selection of teaching as a career.

The largest number of significant different code type frequencies occurred in the group of students who dropped out of the University prior to starting the junior year. In considering this finding the reader should keep in mind that while a sizeable number of academic females (N = 164) was involved, the number of academic males was only 35. Scale 9 (Ma) coded high and in combination with low coded scales was characteristic of the drop-out group. Many of the significant differences occur in patterns involving scale 7 (Pt), especially when scale 7 is coded high with 5 (Mf) coded low. The male profiles with no scales (X) coded high appears to distinguish the two teaching certificate groups from the drop-outs. More comparisons may be possible through further research.

An examination of the data also shows that more differences in frequencies of code types in male groups occurred than in female groups. Possibly there are several reasons for this. Many counselors have found that college females are not as concerned about their academic careers as are male students. Also, the number of students in the female group was larger than the male group in the groups of graduates with teaching certificates.

SECTION IV

Summary and Conclusions

Statement of the Problem

The purpose of this study was to identify and compare profile patterns of teachers with the profile patterns of selected non-teacher groups of students at the University of Wisconsin on the group form of the MMPI. The three main groups studied, divided according to sex, were students who were graduated from the University of Wisconsin with teaching certificates, students with majors in Letters and Science who did not prepare for teaching, and students who dropped out of the University prior to starting the junior year. Those who prepared for teaching were further subdivided into academic and special field majors.

Conclusions were drawn with reference to four questions:

1. What, if any, MMPI profile patterns are found for students who were graduated from the University of Wisconsin with teaching certificates?
   a. Graduates of the School of Education with academic majors.
   b. Graduates with teaching certificates in special fields—Home Economics, Music, Agriculture, Art Education, and Physical Education.

2. What, if any, profile patterns are found for students who were graduated from the College of Letters and Science but did not prepare for teaching?

3. What, if any, profile patterns are found for

students who dropped out of the University prior to starting the junior year?

4. If profile patterns are found, how do they compare among the groups?

While most of the earlier researches on the MMPI were done in a hospital setting, research has been conducted in a variety of situations with normal persons, especially with university students.

The Design of the Study

The study here reported was of the comparative-descriptive type in which groups of prospective teachers and non-teaching groups were compared upon scales and profile patterns of the MMPI. In an exploratory study such as this it seemed best to work with a well defined population about whom precise and rather complete information could be had. The study concerned students enrolled at the University of Wisconsin during the fall semesters of 1949, 1950, and 1951.

The present study was designed to obtain more information about the suitability of using the MMPI with University of Wisconsin students in academic and vocational guidance, especially in the selection of future teachers. The MMPI is a widely used instrument and has been found useful in clinical studies. It is important that its value for this particular purpose should be ascertained.

The data were analyzed through the use of Yule's formula for differentiation of groups.

$$e_{12}2 = \frac{p_1 q_1}{n^1} + \frac{p_2 q_2}{n^2}$$

The significance of differences between the groups were determined by the use of this formula.

Difference in profile patterns were found in six comparisons made for each sex.

Findings

The detailed findings briefly summarized were:

1. The most important and highly significant finding is the fact that statistically significant differences did occur among all groups.

2. It appears from the data that there were fewer differences among the graduates with teaching certificates in the academic fields as compared with the graduates with teaching certificates in the special fields than between any other groups.

3. The largest number of significant code type frequencies occurred in the group of students who dropped out of the University.

4. More differences in frequencies of code types in male groups occur than in female groups.

Conclusions

This study seems to indicate that significant differences in profile patterns on a personality test can be found between groups of college students with different academic histories. While the study has been limited by the small numbers of persons in some sub-groups, nevertheless, personality differences as represented by significant differences in patterns of scales of the MMPI did occur.

Research is being continued to determine what the profile patterns may mean. In this respect, counseling experience and further research may help.

There are many interesting and difficult problems in research design for this area. Group research techniques may be employed in building a usable formula or to make useful generalizations. Our ultimate concern, however, is with individuals.

Further research is needed in this area to determine what the profile patterns may mean and to find a measure that predicts early and adequately the personality variables involved in later success.

BIBLIOGRAPHY

1. Barr, A. S. "Problems Associated with the Measurement and Prediction of Teacher Success," Journal of Educational Research, LI (May 1958), pp. 695-99.
2. Barr, A. S. "The Measurement and Prediction of Teacher Efficiency: A Summary of Investigations," Journal of Experimental Education, XVI (June 1948), pp. 203-83.
3. Barr, A. S., et al. "The Measurement and Prediction of Teaching Efficiency," Review of Educational Research, XXV (1955), pp. 261-69.
4. Barr, A. S., and Lois Douglas. "The Pre-Training Selection of Teachers," Journal of Educational Research, XXVIII (October 1934), pp. 100-6.
5. Blum, L. P. "A Comparative Study of Students Preparing for Five Selected Professions Including Teaching," Teacher Personnel Research, Monograph No. 3 (Madison, Wis.: Dembar Publications, 1947), pp. 31-65.
6. Carlile, A. B. "Predicting Performance in the Teaching Profession," Journal of Educational Research, XLVII (May 1954), pp. 641-68.
7. Cronbach, Lee J. Essentials of Psychological Testing (New York: Harper and Brothers

1949), pp. 305-67.
8. Drake, L. E. "MMPI Profiles and Interview Behavior," Journal of Counseling Psychology, III (1954), pp. 92-5.
9. Drake, L. E. "Interpretation of MMPI Profiles in Counseling Male Clients," Journal of Counselling Psychology, III (1956), pp. 83-8.
10. Drake, L. E., and Oetting, E. R. "An MMPI Pattern and a Suppressor Variable Predictive of Academic Achievement," Journal of Counseling Psychology, IV (1957), pp. 245-47.
11. Dressel, P. L., and Mann, W. "Appraisal of the Individual," Review of Educational Research, XXI (1951), pp. 115-31.
12. Good, Carter V., Barr, A. S., and Scates, D. E. The Methodology of Educational Research (New York: D. Appleton-Century Co., 1938), pp. 648-78.
13. Harmon, L. R., and Wiener, D. N. "Use of the Minnesota Multiphasic Personality Inventory in Vocational Advisement," Journal of Applied Psychology, XXIX (1945), pp. 131-41.
14. Hathaway, S. R. "A Coding System for MMPI Profiles," Journal of Counseling Psychology, XI (1947), pp. 334-37.
15. Hathaway, S. R., and Monachesi, E. D. Analyzing and Predicting Juvenile Delinquency with the MMPI (Minneapolis: University of Minnesota Press, 1953), pp. 13-23.
16. Hawkes, G. R. "Use of the Minnesota Multiphasic Personality Inventory in Screening College Students for Counseling Purposes," Journal of Educational Psychology, XLI (February 1950), pp. 116-21.
17. Klausmeier, H. J. Teaching in the Secondary School (New York: Harper and Brothers, 1958).
18. Klausmeier, H. J. Desirable Characteristics of the High School Teacher of Psychology, paper presented to Symposium, American Psychological Association, 1958.
19. Klausmeier, H. J., and Swanson, D. J. "The Development of a Functional Course in Educational Psychology for Teachers," Journal of Educational Psychology, XLI (1950), pp. 449-72.
20. McDaniel, Henry B. Guidance in the Modern School (New York: Dryden Press, 1950).
21. Meehl, P. E. Clinical versus Statistical Prediction (Minneapolis: University of Minnesota Press, 1954), pp. 17-22.
22. Michaelis, John U. "The Prediction of Success in Student Teaching from Personality and Attitude Inventories," University of California Publications in Education, XI (1954), pp. 415-81.
23. Ringness, T. A. "Relationship Between Certain Attitudes Towards Teaching and Teaching Success," Journal of Experimental Education, XXI (September 1952), pp. 1-55.
24. Welsh, G. S., and Dahlstrom, W. G., Editors, Basic Readings on the MMPI in Psychology and Medicine (Minneapolis: University of Minnesota Press, 1956).
25. Yule, G. U. An Introduction to the Theory of Statistics, 8th edition, revised (London: Griffin and Co., 1927), pp. 254-72.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 4, June 1961)

A COMPARISON OF THE READABILITY OF VOCATIONAL AGRICULTURE REFERENCE BOOKS WITH THE READING ABILITY OF THE STUDENTS USING THEM*

RALPH R. BENTLEY and R. EDWARD GALLOWAY**
Purdue University

Introduction

TEACHERS OF vocational agriculture often complain that their students do not understand what they read. In vocational agriculture, as in other high-school subjects, reading is one of the most effective and efficient tools of learning. However, the usefulness of this tool is lost if the material to be read does not reasonably match the capability of the reader. If the material is too difficult, the student will not comprehend; if it is too simple, the student may be insulted and/or bored. In either case, learning suffers.

Few studies have been reported which attempted to match the readability of text or reference materials with the reading ability of high-school students. Although the readability of a large number of agricultural reference materials has been reported, no extensive study was found which compared their readability with the tested reading ability of the students actually using them.

Purpose

It was the purpose of this study to compare the readability of widely used agricultural reference books with the reading ability of the vocational agriculture students using them. In order to make this comparison, it was necessary to measure both the readability of vocational agriculture reference books and the reading ability of students of vocational agriculture who were using them.

Procedure

Selection of Cooperating Schools. One relatively large high school with an enrollment of over 100 students and one relatively small high school with an enrollment of under 100 students were chosen in each of the twelve vocational agriculture districts in Indiana. Large and small schools were selected in each district on the basis of three criteria: 1) the school had a department of vocational agriculture approved by the Indiana State Department of Public Instruction, 2) the teacher of vocational agriculture had taught at least one year in the school, and 3) both the teacher of vocational agriculture and his administrator expressed their willingness to cooperate in the study. The basic data for this study were obtained by testing the reading ability of the vocational agriculture students in these schools, and by measuring the readability of the agricultural reference books in their libraries.

Measurement of Readability. Responses of the cooperating teachers of vocational agriculture to a survey indicated that almost 100 reference books were available in their libraries in quantities large enough for class use. Ten of the books found most frequently in the cooperating schools were chosen for readability measurement. These books were A) Your Farming Program, B) Shopwork on the Farm, C) Profitable Farm Management, D) Beef Production, E) Livestock and Poultry Production, F) Approved Practices in Beef Production, G) Veterinary Guide for Farmers, H) Crop Production, I) Productive Soils, and J) Feeds and Feeding (abr.).

The Dale-Chall Readability Formula was used to measure the readability or reading difficulty of these ten books. This measurement yielded a raw score for each of twenty or more sample portions of reading material from each book. The readability of each book was expressed as a mean and a standard deviation of the sample scores. Analysis of variance was used to test differences

* Adapted from a Ph. D. dissertation completed at Purdue University.

**The authors wish to express their gratitude to the Purdue Research Foundation for its support of this study.

among the ten books.

Measurement of Reading Ability. Two tests were used to measure the reading ability of vocational agriculture students. The Cooperative English Tests, Test C1, "Reading Comprehension", was used to measure general reading ability, and the Iowa Tests of Educational Development, Test 6, "Ability to Interpret Reading Materials in the Natural Sciences", was chosen as the most appropriate test of ability to read technical materials in agriculture. The Verbal Battery and the Nonverbal Battery of the Lorge-Thorndike Intelligence Tests series were used to determine the mental ability of these students. These tests were administered by the teacher of vocational agriculture in each high school during his regular vocational agriculture class periods. After the scores for each grade from each of the tests were analyzed by analysis of variance, the means and standard deviations were calculated for each school and/or groups of non-differing schools, depending upon the results of the significance tests. These means and standard deviations were converted to grade level units.

Results

The results of this study, designed to compare the readability of certain agricultural reference books with the reading ability of the students using them, are presented under three headings: 1) readability of agricultural reference books, 2) reading ability of vocational agriculture students, and 3) comparison of book readability with student reading ability. In each of these three sections, hypotheses are stated, an analysis of relevant data is presented, and conclusions are drawn. The .05 level of significance was used.

Readability of Vocational Agriculture Reference Books. The raw Dale-Chall scores for sample portions of reading material in each of the ten selected books furnished the basic readability data. Bartlett's test was used to test the homogeneity of within-book variances and a one-way analysis of variance was used to test the significance of differences among the mean Dale-Chall readability scores for the ten books. These tests indicated that significant differences existed both among the variances and the means for the ten books. Therefore, each book was treated individually in the comparison with student reading ability.

Both the mean and the standard deviation readability for each book was converted from raw Dale-Chall score units to grade-level units. This made a direct comparison of readability and reading ability scores possible.

Reading Ability of Vocational Agriculture Students. The raw scores for both the Cooperative Reading Test and the Iowa Reading Test furnished basic reading ability data. The data for each of these tests were analyzed by two-way analysis of variance and presented separately for each of the four high-school grades.

It was hypothesized that the reading ability of ninth-grade students did not differ significantly with respect to the size and location of the schools attended. The results of the analysis of variance for both of the reading tests and the significance of the obtained F values are shown in Table I.

The reading ability of ninth-grade students as measured by the Cooperative and Iowa Reading Tests (Table I) did not differ significantly among the cooperating schools when they were grouped by district and size. Thus it was concluded that the differences among districts and between sizes could have occurred by chance and that these students belong to the same population with respect to reading ability as measured by these two tests. Population means and standard deviations for the Cooperative and the Iowa Reading Tests are shown in Table II.

It was hypothesized that the reading ability of tenth-grade students did not differ significantly with respect to the size and location of the schools attended. The results of the analysis of variance for each test and the significance of the obtained F values are shown in Table III.

The reading ability of tenth-grade students as measured by the Cooperative Reading Test did not differ significantly among the cooperating schools when they were grouped by district and size (Table III). Thus it was concluded that the differences among districts and between sizes could have occurred by chance and that these tenth-grade students of vocational agriculture belong to the same population with respect to their reading ability as measured by the Cooperative test. Table III also shows that ninth-grade students are significantly different from district to district on reading ability as measured by the Iowa test.

The significant F values for interaction in Table III indicate that unique effects of school size occur in some districts for reading ability as measured by the Iowa test. Therefore, it was concluded that students in the various districts belong to different populations and that students in individual schools belong to different populations with respect to their reading ability as measured by the Iowa test. The population mean and the standard deviation for the Cooperative Reading Test are shown in Table IV. The means and standard deviations for the Iowa Reading Test are given in Table V for students in each large and small school by districts.

It was hypothesized that the reading ability of eleventh-grade students of vocational agriculture did not differ significantly with respect to the size and location of the schools they attended. The results of the analysis of variance for both reading tests and the significance of the obtained F values are shown in Table VI.

TABLE I

TWO-WAY ANALYSIS OF VARIANCE OF READING SCORES MADE BY NINTH-GRADE STUDENTS OF VOCATIONAL AGRICULTURE

Source of Variance	Weighted Sums of Squares	d. f.	Mean Square	Significance of F Value*
		Cooperative Reading Test		
District	1640. 4132	10	164. 0413	N. S.
Size	21. 3928	1	21. 3928	N. S.
Interaction	1168. 760	10	116. 8760	N. S.
Within	20782. 239	194	107. 1249	
		Iowa Reading Test		
District	389. 4164	10	38. 9416	N. S.
Size	58. 1405	1	58. 1405	N. S.
Interaction	258. 219	10	25. 822	N. S.
Within	4425. 575	191	23. 1705	

*N. S. indicates that the observed F was not significant at the . 05 level.

TABLE II

MEANS AND STANDARD DEVIATIONS FOR THE COOPERATIVE AND IOWA READING TESTS FOR NINTH-GRADE STUDENTS IN TWENTY-TWO SCHOOLS

Tests	Mean Grade Level of Ability	Grade Levels of S. D.
Cooperative Reading	7. 5	2. 5
Iowa Reading	8. 7	2. 0

TABLE III

TWO-WAY ANALYSIS OF VARIANCE OF READING SCORES MADE BY TENTH-GRADE STUDENTS OF VOCATIONAL AGRICULTURE

Source of Variance	Weighted Sums of Squares	d. f.	Mean Square	Significance of F Value*
		Cooperative Reading Test		
District	2561.5332	10	256.1533	N.S.
Size	191.1195	1	191.1195	N.S.
Interaction	2274.980	10	227.498	N.S.
Within	24476.388	176	139.0704	
		Iowa Reading Test		
District	482.5512	10	48.2551	$P < .05$
Size	.010043	1	.010043	N.S.
Interaction	641.842	10	64.1842	$P < .01$
Within	4134.81	174	23.7633	

*N.S. indicates that the observed F was not significant at the .05 level.

TABLE IV

MEAN AND STANDARD DEVIATION FOR THE COOPERATIVE READING TEST FOR TENTH-GRADE STUDENTS IN TWENTY-TWO SCHOOLS

Test	Mean Grade Level	Grade Levels of Standard Deviation
Cooperative Reading	8.3	3.3

TABLE V

MEANS AND STANDARD DEVIATIONS FOR THE IOWA READING TEST FOR TENTH-GRADE STUDENTS BY DISTRICT AND SCHOOL SIZE

District	Large		Small	
	Mean Grade Level	Standard Deviation	Mean Grade Level	Standard Deviation
	Iowa Reading			
I	10.0	2.8	7.6	.9
II	9.4	1.2	11.1	4.2
III	9.8	2.3	11.9	4.6
IV	11.1	4.9	10.4	2.8
V	8.6	1.0	10.8	3.2
VI	11.1	3.8	12.4	5.3
VII	11.1	3.0	8.0	0.3
VIII	9.8	2.3	9.0	2.3
IX	8.6	1.8	9.7	1.6
XI	9.3	3.1	12.2	3.7
XII	9.6	2.9	8.5	2.7

The reading ability of eleventh-grade students as measured by the Cooperative and Iowa tests did not differ significantly among the twenty-two cooperating schools when they were grouped by district and size. Thus it was concluded that the differences observed among districts and between sizes could have occurred by chance and that these students belong to the same population with respect to their abilities as measured by these two tests.

Population means and standard deviations for the Cooperative and the Iowa Reading Tests are shown in Table VII.

It was hypothesized that the reading ability of twelfth-grade students did not differ significantly with respect to the size and location of the schools they attended. The results of the analysis of variance for each of the reading tests and the significance of the obtained F values are shown in Table VIII.

The reading ability of twelfth-grade students as measured by the Cooperative and the Iowa Reading Tests did not differ significantly among districts but did differ significantly between large and small schools. Therefore, it was concluded that the differences between large and small schools probably did not occur by chance and that the twelfth-grade students in large schools belong to a different population with respect to the Cooperative and the Iowa Reading Tests. Means and standard deviations are shown in Table IX for twelfth-grade students on the Cooperative and Iowa Reading Tests by large and small schools.

Comparison of Book Readability with Student Reading Ability. In order to determine how appropriate vocational agriculture reference books are for the students who use them, comparisons were made on the basis of the readability grade levels of reference books as determined by the Dale-Chall formula and the grade levels of reading ability of students as measured by the Cooperative and Iowa Reading Tests. Both mean and standard deviation grade levels were used in these comparisons. Figure 1 shows these means and standard deviations graphically.

The upper portion of Figure 1 presents the data regarding ten agricultural reference books. Each bar in this portion of the figure shows the mean and the range, minus one to plus one standard deviation, in grade levels of book readability. For example, the first bar in Figure 1 shows that the book, Your Farming Program, has eighth-grade mean readability and that two-thirds of the reading material in this book falls between 6. 4 and 9. 6 grade levels.

The lower portion of Figure 1 graphically presents information that was shown in Tables II, IV, V, VII, and IX about the reading ability of students. Each bar in this portion of the figure shows the mean and the range in grade levels of student reading ability from minus one to plus one standard deviation. Thus, the ninth-grade students of vocational agriculture in the twenty-two cooperating schools had mean general reading ability of grade 7. 5 as measured by the Cooperative Reading Test and two-thirds of these students had a reading ability range of grade 5. 0 and 10. 0.

It will be noted that the mean reading ability of students for each grade level was higher when measured by the Iowa Reading Test than when measured by the Cooperative Reading Test. It will also be noted that mean student reading ability was higher for each succeeding high-school grade except for twelfth-grade students in the large high schools. However, reading ability, as measured by these tests, fails to increase a full grade level for students in each succeeding high-school grade.

Since there was significant interaction among schools with respect to the Iowa Reading Test scores for tenth-grade students, it was not permissable to pool these scores for all tenth-grade students. To show the extreme range of reading ability of tenth-grade students, the highest and the lowest means and their respective standard deviations are shown in Figure 1.

In general, Figure 1 shows that the means for book readability are higher than the means for student reading ability. The grade level means for the ten books studied range from grades eight to eleven with standard deviations ranging from 1. 6 to 2. 7 grade levels. The student grade level means for scores on the Cooperative Reading Test range from 7. 5 to 9. 9 with standard deviations ranging from 2. 5 to 3. 3.

In the following paragraphs the readability of the books as measured by the Dale-Chall formula will be compared with the reading ability of vocational agriculture students as measured by the Cooperative and the Iowa Reading Tests. Both readability and reading ability will be expressed as grade levels. For example, a pupil with ninth-grade reading ability as measured by the Cooperative Reading Test has a score equal to the median score made on this test by students in the publisher's norm group who were halfway through their ninth year of school. Ninth-grade readability implies that the reading material can be read by students who are halfway through the ninth grade. These comparisons are presented separately for each of the four high-school grades.

Figure 1 shows that all ten of the books studied have mean readability scores above the mean reading ability for ninth-grade students as measured by the Cooperative Reading Test. Although these same students scored a grade higher on the Iowa Reading Test, nine of the ten books are still, on the average, too difficult for these ninth-grade students.

The overlapping of standard deviation bars in in Figure 1 shows that while the most able ninth-

TABLE VI

TWO-WAY ANALYSIS OF VARIANCE OF READING SCORES MADE BY ELEVENTH-GRADE STUDENTS OF VOCATIONAL AGRICULTURE

Source of Variance	Weighted Sums of Squares	d. f.	Mean Square	Significance of F Value*
		Cooperative Reading Test		
District	1366.0296	10	136.6030	N. S.
Size	254.10000	1	254.10000	N. S.
Interaction	1946.477	10	194.6477	N. S.
Within	16080.218	156	103.0783	
		Iowa Reading Test		
District	69.6304	10	6.9630	N. S.
Size	47.3594	1	47.3594	N. S.
Interaction	388.373	10	38.8373	N. S.
Within	3777.077	158	23.9056	

*N. S. indicates that the observed F was not significant at the .05 level.

TABLE VII

MEANS AND STANDARD DEVIATIONS FOR THE COOPERATIVE AND IOWA READING TESTS FOR ELEVENTH-GRADE STUDENTS IN TWENTY-TWO SCHOOLS

Tests	Mean Grade Level	Grade Levels of Standard Deviation
Cooperative Reading	9.1	2.6
Iowa Reading	10.0	3.0

TABLE VIII

TWO-WAY ANALYSIS OF VARIANCE OF READING SCORES MADE BY TWELFTH-GRADE STUDENTS OF VOCATIONAL AGRICULTURE

Source of Variance	Weighted Sums of Squares	d. f.	Mean Square	Significance of F Value*
		Cooperative Reading Test		
District	1127. 5896	10	112. 2590	N. S.
Size	880. 5533	1	880. 5533	$P < .01$
Interaction	1015. 1126	10	101. 5113	N. S.
Within	14384. 024	132	108. 9699	
		Iowa Reading Test		
District	423. 3408	10	42. 3341	N. S.
Size	138. 4966	1	138. 4966	$P < .05$
Interaction	232. 851	10	23. 2851	N. S.
Within	4422. 243	132	33. 5018	

*N. S. indicates that the observed F was not significant at the . 05 level.

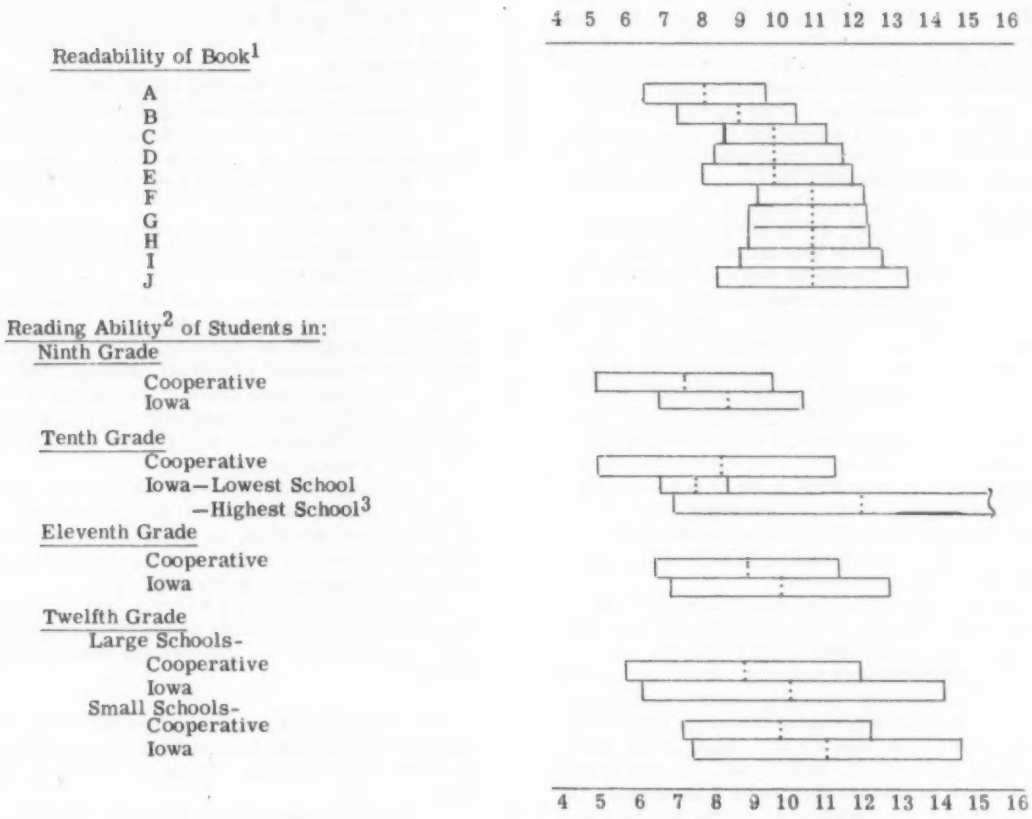
TABLE IX

MEANS AND STANDARD DEVIATIONS FOR THE COOPERATIVE AND IOWA READING TESTS FOR TWELFTH-GRADE STUDENTS SHOWN BY LARGE AND SMALL SCHOOLS

	Large		Small	
Tests	Mean Grade Level	Grade Level of S. D.	Mean Grade Level	Grade Level of S. D.
Cooperative Reading	9. 0	3. 3	9. 9	2. 7
Iowa Reading	10. 3	4. 2	11. 2	3. 7

FIGURE I

GRAPHIC ILLUSTRATION OF MEANS AND STANDARD DEVIATIONS FOR DALE-CHALL READABILITY OF TEN AGRICULTURAL REFERENCE BOOKS AND FOR READING ABILITY OF VOCATIONAL AGRICULTURE STUDENTS AS MEASURED BY THE COOPERATIVE AND IOWA READING TESTS



[1]Titles of these ten books are found in the list of books at the end of this study.
[2]Reading ability as measured by the Cooperative and the Iowa reading tests previously described.
[3]Since no pooling was permissable here, only the extreme schools are shown.

grade readers can be expected to read all but the most difficult portions of these ten books, the least able readers could be expected to read only the easiest portions of books A and B.

It is shown in Figure 1 that nine of the ten books have mean readability scores higher than the mean reading ability scores for the tenth-grade students as measured by the Cooperative Reading Test. When compared to mean reading ability as measured by the Iowa Reading Test, the average reading ability for each of the ten books exceeds the lowest tenth-grade class mean and is exceeded by the highest tenth-grade class mean.

The standard deviation bars show that the most able tenth-grade students can be expected to read a major part of the material in each of these ten books and that all but the poorest readers can be expected to read selected portions of these books.

Figure 1 shows that the mean readability of two of the ten books was below the average reading ability of eleventh-grade students and of twelfth-grade students in large schools as measured by the Cooperative Reading Test.

Eight books were above their reading ability. When reading ability was measured by the Iowa test, only half of the books were too difficult, and two were somewhat too easy for these students.

The standard deviation bars indicate that the more able readers in the eleventh grade and in the twelfth grade of large schools can be expected to read all except the most difficult portions of these ten books and that these students will possibly find much material in books A through E too easy. The poorest eleventh-grade readers can be expected to read most of books A and B and the easier portions of the other eight books, while the least able twelfth-grade readers in large schools will probably experience difficulty in reading any of these books.

Figure 1 shows that the twelfth-grade students in the small schools demonstrate reading ability about one grade level higher than the twelfth-grade students in large schools. When compared on their Cooperative Reading Scores, the ability of these students was above the mean readability of books A and B; was about equal to the readability of books C, D, and E; and was one grade level below the means for books F through J. On the basis of their Iowa scores, these students had average ability slightly above the mean readability for each of the ten books.

When the standard deviation bars for the books were compared with those for the twelfth-grade students in small schools, Figure 1 shows that the most capable readers in this group can be expected to read practically all of the material in the ten books with ease. However, these books contain few portions which the least able readers can read.

Conclusions and Implications

The following conclusions were drawn on the basis of reading ability data obtained from the vocational agriculture students in cooperating schools and readability data from reference books selected from the libraries of those schools. In the interpretation of these conclusions, one should be mindful of the limitations imposed by the design and measurements used in this study.

1. Each of the vocational agriculture reference books had a mean readability appropriate for students of average reading ability in one of the four high-school grades.
2. Twelfth-grade students, especially those in large schools, were characterized by their low reading ability.
3. In general, reading ability did not differ with respect to school size and location.
4. In general, verbal mental ability did not differ with respect to school size and location.
5. Nonverbal mental ability varied widely among individual schools.
6. Reading ability varied widely both among vocational agriculture classes and within classes at a given grade level.
7. Students of vocational agriculture scored from one to two grade levels higher on the Iowa Test of Reading in the Natural Sciences than they did on the Cooperative General Reading Test.
8. Vocational agriculture students had mean reading abilities which ranged from zero to three grade levels below their peers on the basis of publisher's norm groups.
9. Although there was wide variation from school to school in the mental ability of vocational agriculture students, in general, their mental ability was similar to that of their peers as represented by publisher's norm groups.
10. In general, the agricultural reference books used by the students of vocational agriculture tended to be too difficult for their reading ability.

The findings and conclusions of this study appear to have several implications for workers in the field of education. These are listed as follows:

1. It seems apparent that if a single vocational agriculture textbook were selected for use by students in a given grade, some students would experience difficulty reading any but the most readable portions and some would find all but the most difficult portions distressingly easy. Therefore, it seems desirable to use a variety of reference materials with varied readability in each agricultural subject area.
2. Teachers cannot assume that a student is an average reader because he possesses average

mental ability.

3. When authors and publishers prepare agricultural reference books, they should give special attention to book readability and the reading ability of vocational agriculture students.

4. The fact that vocational agriculture students consistently scored higher on the test of ability to read natural science materials than on the general reading tests suggests that the ability to read agricultural references may be specific and different from general reading ability.

5. Teachers of vocational agriculture should include readability as a factor in the selection of reference books.

6. Teachers of vocational agriculture should give special attention to the reading problems encountered by their students.

TITLES OF AGRICULTURE REFERENCE BOOKS USED IN THIS STUDY

A. Carsie Hammonds, and W. R. Tabb, Your Farming Program (Lexington, Ky.: Hurst Printing Co., 1951).

B. Mark M. Jones, ShopWork on the Farm (New York: McGraw-Hill Book Co., 1945).

C. James E. Hamilton and W. R. Bryant, Profitable Farm Management (New York: Prentice Hall, 1956).

D. Ronald V. Diggins and Clarence E. Bundy, Beef Production (New York: Prentice Hall, 1953).

E. Clarence E. Bundy, Livestock and Poultry Production (New York: Prentice Hall, 1953).

F. Elwood M. Juergenson, Approved Practices in Beef Production (Danville, Ill.: The Interstate, 1958).

G. G. W. Stamm, Veterinary Guide for Farmers (Chicago: Windsor Press, 1950).

H. Henry L. Algren and Richard J. Delorit, Crop Production (New York: Prentice Hall, 1953).

I. Wilbert Walter Weir, Productive Soils (Chicago: J. B. Lippincott Co., 1946).

J. Frank B. Morrison, Feeds and Feeding, abridged (Ithaca, New York: The Morrison Publishing Co., 1958).

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 4, June 1961)

THE RELATIONSHIP OF CERTAIN LETTER FORM VARIANTS IN THE HANDWRITING OF FEMALE SUBJECTS TO THEIR EDUCATION, I.Q. AND AGE

LAWRENCE EPSTEIN, HUNTINGTON HARTFORD
IRVING TUMARKIN*
Columbia University
College of Physicians and Surgeons
Handwriting Institute, Inc. **

THE MODEL of good penmanship has remained the same throughout the United States for many years, despite the fact that the methods used to teach handwriting have undergone considerable change. As a result, this basic handwriting model, which has been called the "school-copy" model by Lewinson and Zubin (6), may be used as a yardstick to evaluate the tendency of individuals to deviate from what they first learned.

Everyone who writes has deviated somewhat from the "school-copy" form by altering the forms and size ratios of his letters in an idiosyncratic way. It is for this reason that signatures can be accepted as legal identification all over the world. Furthermore, it is the fact that each individual learns to write in a unique but consistent way which underlies the science of graphology.

Two essential reasons that each individual departs from the "school-copy" form are to achieve efficiency and speed. Naturally, the demand for legibility remains even after the demand for rigid reconstruction of the school-copy form is relaxed. Typically, handwriting is not taught to students over 14 years of age and thus, it is then that the tendency to vary accelerates. Many students crystallize their own writing styles in high school or, perhaps, in college when they are called upon to take lecture notes at high speeds.

If handwriting typically departs from the school-copy model, what then is reflected in those adult handwritings which closely adhere to, rather than depart from, the school-copy model? If we regard such samples as cases of arrested handwriting development it would seem to follow that such handwritings may also reflect an arrested intellectual and/or emotional development.

In a recent study, Epstein and Hartford (4) found that in a group of male medical students the retention of the superfluous initial upstrokes on the letters u, v, w, and y[1] was, in fact, related to emotional immaturity. In another recent unpublished study, Linton, Epstein and Hartford found the use of these initial upstrokes to be associated "with relative inadequacy and lack of energy in intellectual functioning and other forms of coping with the environment..."

In our survey of earlier graphological studies, we can find but few studies dealing with the relationship between handwriting and intelligence in adult subjects.

Thorndike (11) and Omwake (7) have reported virtually zero correlations between scales of handwriting quality[2] and adult intelligence. These results, in effect, indicate the lack of relationship between adult intelligence and conformity to such school-copy values as legibility, aesthetic appearance, and neatness. In children, however, since they learn to write while under strong external pressure to conform to the school-copy model, correlations of about .30 have been found by Gesell (5) and Starch (10) between ratings of handwriting quality and general scholarship.

The Thorndike and Omwake studies, in having established that adherence to school-copy values in adult handwritings is not positively related to intelligence, raised the further question of what departures, if any, from the school-copy might be associated with intelligence in adult subjects. A study performed by Castelnuovo-Tedesco (3) points the way. He instructed six untrained judges to rate 100 adult handwritings for intelligence and originality, separately, and then correlated these ratings with I.Q. scores. As criteria for intelligence, the judges were told to use aesthetic spacing, handwriting speed, and the simplification (and/or) deviations of the letter forms. They were told to use as the main criteria for judging originality "the extent to which the writer of a

*All footnotes will be found at end of article.

TABLE I

g, d, AND f VARIANTS AND THE FREQUENCY AT WHICH THEY OCCUR AT LEAST ONCE IN A WRITING SAMPLE
(N = 150)

g		d		f	
Variant	f	Variant	f	Variant	f
	13		6		21
	19		7		21
	34		25		31
	50		66		43
	60		129		66
	74		136		95
	143				137

*School copy

given script has followed, in the process of writing a specific letter, a different-writing-pattern from that which he was probably taught in school." These correlations were all at least .50, and since 100 samples were rated it is extremely unlikely that the obtained relationships came about as a result of sampling errors.

The only other study in which judges were successful in predicting handwriting from intelligence is that reported by Binet (2). He used professional graphologists. Unfortunately, however, the handwriting criteria upon which these graphologists based their predictions have not been indicated.

In the present study, our purpose was to explore the relationship of particular deviations from the school-copy letter forms to the educational level, the I.Q. and the age of a group of female writers.

Method

Samples of the writing of 150 female clients were selected from the files of the NYU Testing and Advisement Center. An examination of them suggested three letters as worthy objects of investigation because they permitted wide graphological variations within the three handwriting zones—upper, middle and lower, without loss of identity. These letters were g, d, and f. The more predominant script variations of these letters within our sample were then noted.

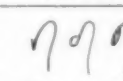
In addition to the school-copy form, there were five form variants of the letter d, six form variants of the letter g, and six of the letter f. For purposes of analysis, the appearance of at least one instance of a particular letter variant in a writing sample was considered sufficient to describe that variant as present in this sample. For each variant of the letter g, d, and f, the number of samples in which it was present was determined. For instance, the school-copy g appeared at least once in 143 of the 150 writing samples. Table I presents the form variants for each of the letters and the number of Ss who used each variant at least once.

The experimenters also were struck by an additional handwriting phenomenon. Some Ss economized in a special way in writing the word "of". They would convert the upstroke normally used to round off the letter o directly into the initial upstroke of the following f. We selected this phenomenon for separate study and it is shown in Table II, which also presents the number of samples in which it was present at least once.

The ages, I.Q.'s and the highest grades attained by the subjects were recorded. The ages ranged from 15 to 55 and the median age was 19 years. The I.Q. scores, which were either scores on the Wechsler Bellevue Intelligence Scale or the Wechsler Adult Intelligence Scale, ranged from 80 to 145. The median I.Q. score was between 124 and 125. It was decided to use 125 as a cut-off point for various statistical tests.

TABLE II

"Of" CONNECTIVE AND THE FREQUENCY AT WHICH IT OCCURS AT LEAST ONCE IN A WRITING SAMPLE
(N = 150)

Variant	f
	45

As for educational level, the subjects at the lowest level had completed the eleventh grade, and the subjects at the highest level had completed more than four years of college (Table III). The subjects were divided into two groups on the basis of grade completed, one group consisted of those who had gone no further than the twelfth grade (the first two groups on Table III), and the other consisted of those who had gone to college for at least some period.

In sum, the distributions of scores on three variables, age, I.Q. and grade completed, were dichotomized for purposes of doing statistical tests. The best approximation of the median was used as the cut-off point in each case.

Characteristics of the Sample

A preliminary analysis of the criterion variable was made to throw light on the nature of the sample. The correlation between age and education was .42. The vast majority of subjects were still in school so that this correlation was to be expected, since educational level depends upon age among subjects who are of school age.

The correlation between age and I.Q. was .27. The older subjects in the sample were predominantly those who had continued their education beyond the period of compulsory attendance. Undoubtedly such subjects are mainly from the upper intelligence range, so there was a selective factor in favor of high I.Q. scores at the upper age levels. The correlation between education and I.Q. was .34 for the same reason.

Results and Discussion

The relationship of the d variant to age, education and I.Q. are presented in Tables IVa and IVb. Table IVa, containing the phi coefficients for the more frequently appearing d variants, shows only one significant negative relationship, a negative one between age and the third d variant. (Signi-

TABLE III

DISTRIBUTION OF THE HIGHEST GRADES COMPLETED BY THE 150 FEMALE SUBJECTS

Highest Grade Attained	Number of Subjects
Eleventh grade or lower	26
Twelfth grade	59
Some college	35
Four years of college or more	30

TABLE IVa

CORRELATION COEFFICIENTS OF d VARIANTS WITH EDUCATION, I.Q., AND AGE

Variant	Education	I.Q.	Age
	.12	.03	-.05
	.04	-.12	.10
	.09	.05	-.19*
	.08	.04	-.13

*Significant at .05 or above.

TABLE IVb

CHANCE PROBABILITY FOR RELATIONSHIPS BETWEEN GREEK d, MODIFIED GREEK d, AND EDUCATION, I.Q., AND AGE*

Variant	Education	I.Q.	Age
	.007	.181	.10
	.005	.356	.220

*The one-tailed level of significance is read directly from the table.

icance is considered established at the .05 level unless otherwise indicated.) Table IVb, containing the exact probability ratios for the two infrequently appearing d variants, shows each of these variants to be positively associated with educational level at better than the .01 level on a one-tailed test of significance. A one-tailed test was used because of the suggestion in the literature of the relationship between these d forms and intellectual ability (8). (The exact probability ratios for age and I.Q. showed a positive relationship between these d forms though this relationship did not reach significance.) Because of its resemblance to the Greek alphabet delta, graphologists have traditionally referred to the d variant presented first in Table IVb as the Greek-d. We shall refer to both of these variants as the Greek-d types.

The phi coefficients for the f and g variants are presented in Table V. The variants are ordered so that the one having the most positive correlation with educational level is listed first and the variant having the strongest negative correlation is listed last. Table V also shows the phi coefficients with I.Q. and age. Incidentally, the phi coefficient reflects the frequency of appearance of the variants as well as the strength of their relationships to other variables, so that comparisons should be made with caution.

Table V shows that there are three f variants which have significant positive associations with educational level, although none are significantly associated with I.Q. Among the g variants, two have significant positive associations both with educational level and I.Q. These are identified as the figure-8 and half-figure-8 g. The figure-8 g also has a significant positive association with age. One g variant is significantly and negatively associated both with age and educational level.

Considerable deviation from the school-copy model is reflected in those d, f, and g variants which were found to be associated with higher educational level, I.Q. or age. The question arises how these variants deviate from the school-copy model. We shall consider these variants, first, with respect to the relative degree of simplification which they have undergone and, second, with respect to the relative degree of motoric economy with which they are written.

Simplification is rated roughly in terms of how much graphite or ink is used to make a particular variant, as compared with the amount used in the school-copy form. Among the variants positively associated with educational level, the Greek-type d's, first and second f variants in Table V and the half-figure-8 g all involve a considerable degree of simplification. Yet, the third f in Table V is among the least simplified f variant, and the figure-8 g, which correlates more highly with educational level, I.Q. and age than any other single variant, is less simplified than the last g in Table V which is negatively associated with age and educational level. It is apparent that simplification alone cannot account for the obtained relationships.

Turning now to economy, in graphology, economy usually refers to the amount of energy and time which is involved in the writing. Hence, a letter form is judged more economical if less time elapses and less energy is utilized between the termination of the preceding letter and the beginning of the following letter. An emphasis on rightward tending strokes, an avoidance of leftward tending strokes, and a minimum of breaks and stoppages are therefore the key criteria.

The d, f, and g variants, with the exception of the second f on Table V, that were found to be significantly associated with higher educational level, I.Q. or age all show a high degree of economy. However, the second d in Table IVa, the fourth and fifth f variants in Table V, and the third and sixth g variants in Table V are also very economical forms but not significantly associated with age, I.Q. or education. Therefore, economy alone also cannot account for the relationships which were obtained.

Beyond having undergone simplification and economization, those d, f, and g variants which were positively associated with higher educational level, I.Q. or age, also represent the most original deviations from the school-copy form. Considering the f variants from this viewpoint, we can see that nothing original has been evolved in any of the three variants following the school-copy one in Table V. The transformations from the school-copy have been mainly in the direction of simple omission—of either, or both upper and lower loops. In each of the three f variants which were positively associated with educational level we see the emergence of a new form relative to the school-copy model.

In the case of the figure-8 and the half-figure-8 forms, the transformation when compared with other g forms seems even greater than in the case of the f. It is interesting to note what the American graphologist, Louise Rice (8) has said concerning the figure-8 g; "People who use this formation. . . have a certain amount of intuition and instinct; they are the people who innately judge literature and art fairly well, even though not trained for such judgments. . . This g also shows innate understanding of humanity."

The upper and middle zones of the Greek-d did not differ in as specific a fashion from their school-copy counterparts as did the zonal elements of the higher f and g forms. Yet, the Greek-d types taken as Gestalten, seem to embody an aesthetic simplicity not found in the other d variants. Of the Greek-d, Louise Rice (8) has said, "It is not surprising to find stylists among writers and poets using this d exclusively, nor to find that college

TABLE V

CORRELATION COEFFICIENTS OF g AND f VARIANTS WITH EDUCATION, I.Q., AND AGE

f				g			
Variant	Education	I.Q.	Age	Variant	Education	I.Q.	Age
	.19*	.01	.13		.33**	.23**	.19*
	.19*	.16	.10		.23**	.19*	.10
	.17*	.13	.08		.05	.05	.13
	.15	.00	.06		-.04	.02	-.04
	.14	-.06	.05		-.09	-.08	-.06
	-.03	-.14	-.15		-.13	.00	.10
	-.11	-.09	-.19*		-.22**	-.10	-.18*

* Significant at .05 level.
**Significant at .01 level.

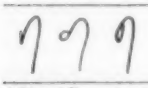
professors and scientists innately take to it. When found in ordinary writing, this formation is the 'straw in the wind' which tells of the intellectual possibilities of the writer. . . ."

Turning now to the d and f variants which were negatively associated with age, and to the g variant which was negatively associated both with age and educational level, we note that none of the three shows any evidence of originality. The f, of course, is the school-copy model; the g variant may be described as an "unfinished" school-copy g, and the d variant is a somewhat deteriorated or inept school-copy d.

Table VI reveals the "of-connective" to have a significant positive association to educational level, I.Q. and age. The direct conversion of the upstroke of the o into the initial upstroke of the f is a radically economical measure which transforms the separate letters into a unified, simplified Gestalt.

TABLE VI

CORRELATION COEFFICIENT OF "of" CONNECTIVE WITH EDUCATION, I.Q., AND AGE

Variant	Education	I.Q.	Age
	.31*	.23*	.35*

*Significant at .01 level.

Among the twenty letter-form variants which we have studied, only the figure-8 g and the half-figure-8 g were found to have a significant, positive association with higher I.Q. The "of-connective"—the only connective form which we included in this study—was also found to be associated with higher I.Q. At this point, we might wonder whether a comparably extensive study of other connective-form variants might yield a greater number of significant positive correlations with I.Q. than we have found among letter-form variants. To the extent that the need to produce a more or less legible script limits one's freedom to forge new pathways, the limitation may be greater on the individual letter forms than on the modes of combining letter forms.

Summary and Conclusions

Each of the small-letter forms selected for study were written at least once in the school-copy manner by about 90 percent of our group of 150 late adolescent and adult female Ss. In addition to the school-copy forms, we were able to classify five variants of letter d, six variants of letter f, and six variants of letter g. These variants appeared in far fewer handwritings than did the school-copy form. Also selected for separate study was a highly economical variation of the word "of".

Among the d, f, and g variants, some adhered rather closely to the school-copy form and some deviated to a much greater extent. We found the school-copy f and one d variant to be significantly associated with lower age, and one g variant to be significantly associated with both lower age and lower educational level. Both the d and f variants departed only slightly from their school-copy counterparts.

Among the seven letter-form variants which were shown to be associated with higher educational level, there were two d variants, three f variants, and two g variants. Both g variants were also associated with higher I.Q. and one, with higher age. All seven of these variants are clearly set off from the remaining eleven variants in that they combine a high degree of motoric economy with the most radical and original departures from the school-copy form. They combine efficiency with creativity and can be considered the most highly evolved of all the variants we have studied. The special case of the of-connective, which is graphologically also a highly developed variation, was also found to be associated with higher educational level, higher I.Q. and higher age.

In effect, our results suggest that the female adult who continues to write in the fashion that would have pleased her elementary school teacher is less likely to be as well educated, as bright, or as mature as the adult who has worked her way out of the school-copy rut and has evolved a more efficient and original way of writing.

The results of the present study, in revealing that there is an evolutionary trend away from the school-copy model, have led us to wonder whether the values underlying the teaching of handwriting should be revamped. Instead of attempting to cast all children in the school-copy mold, perhaps the teacher should be taught to evaluate handwriting progress in the child with respect to the efficiency and legibility of his letter forms as they appear in the context of cursive writing. With this orientation, the teacher may be less likely to make the error of diagnosing spontaneous, creative and economical deviations from the school-copy as an expression of poor progress. Why shouldn't the child who, for one reason or another, can never be a facile copyist, be encouraged to persist in whatever unique modes of writing he discovers, so long as they are no less efficient nor legible than the school copy? It would seem that an early acceptance by teachers of unique styles of writing would make writing more of a pleasure than a chore for the child who is learning to write. More important is the likelihood that many children might be spared repeated experiences of failure to write

"correctly". Such experiences might be the predominant cause of the immature, inept handwritings that are seen in later years—handwritings which persist as foci of self-consciousness, embarrassment, and shame.

FOOTNOTES

* The authors wish to express their appreciation to Dr. Wallace Gobetz of the New York University Testing and Advisement Center, whose cooperation made this study possible.

** The Handwriting Institute, Inc., is a privately supported research organization formed to investigate graphological and graphomotor variables, and to collate the research of others in this area.

1. In learning the school-copy model, we are all taught, when we write words beginning with any one of these four letters, that our first stroke should be an upstroke, originating at the base line. Since this beginning stroke is completely superfluous to the essential letter form, and since it tends to reduce the efficiency and the fluency of the writing, people tend to omit it as they mature.

2. Scales of handwriting quality such as those of Ayres (1), Thorndike (12), Rosenhaus (9), are commonly used in the elementary schools to evaluate the individual child's progress in cursive writing. The scales all typically consist of a graded series of handwriting samples, from worse to better, the best and highest grade sample always being an ideal school-copy model.

REFERENCES

1. Ayres, L. P. A Scale for Measuring the Quality of Handwriting of Children (New York: Russel Sage Foundation, Division of Education, 1917).
2. Binet, A. Les Revelations de l'Ecriture d'apres un Controle Scientifique (Paris: Alcan, 1906).
3. Castelnuovo-Tedesco, P. "A Study of the Relationship Between Handwriting and Personality Variables," Genetic Psychological Monographs, XXXVII (1948), pp. 167-220.
4. Epstein, L. and Hartford, H. "Some Relationships of Beginning Strokes in Handwriting to the Human Figure Drawing Test," Perceptual Motor Skills, IX (1959), pp. 55-62.
5. Gesell, A. "Accuracy in Handwriting as Related to School Intelligence and Sex," American Journal of Psychology, XVII (1906), pp. 394-405.
6. Lewinson, T., and Zubin, J. Handwriting Analysis (New York: King's Crown Press, 1944).
7. Omwake, K. T. "The Value of Photographs and Handwriting in Estimating Intelligence," Public Personnel Studies, III (1925), pp. 2-15.
8. Rice, L. Character Reading from Handwriting (New York: Frederich A. Stokes Co., 1927).
9. Rosenhaus, M., and others. Survey Chart for Cursive Handwriting (New York: Board of Education of City of New York, 1949).
10. Starch, D. "The Measurement of Handwriting," Journal of Educational Psychology, IV (1913), pp. 445-464.
11. Thorndike, E. L. "Handwriting," Teachers College Record, XI, No. 2 (1910).
12. Thorndike, E. L. Handwriting Scale (New York: Bureau of Publications, Teachers College, Columbia University, 1910).

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 4, June 1961)

A STUDY OF PATHOLOGICAL DOUBTING AS A RESPONSE SET

VELMA I. RUST*
University of Illinois

FOR YEARS test makers have been concerned with designing tests which will measure accurately knowledge of content, specific abilities or skills, personality traits, and so on. They are interested in eliminating, or keeping to a minimum, all extraneous effects caused by guessing, or mental or attitudinal sets on the part of the subjects. Validity is kept high by giving special attention to the wording of items, their arrangement, and keying.

A few writers (3, 4) recently have raised questions regarding the relationship between evidence of knowledge of content (in the form of test scores) and a tendency to be very hesitant about accepting any statement as proven. The author became interested in studying the response set herein called pathological doubting. The definition of a pathological doubter used by Ennis (3) is the one also used here. According to him, a pathological doubter is a person who is not just cautious, but who will never accept anything as proven, a person for whom "proof", "sufficient evidence", and "warranted" are inapplicable in all conceivable situations.

The major purpose of this research was to test two hypotheses: 1. pathological doubters, chosen by the use of criteria based on their responses to selected items of the Watson-Glaser Critical Thinking Appraisal, consistently exhibit a pattern of reasoning which differs from that of the other subjects in the sample; and 2. the presence of a response set such as pathological doubting affects the apparent factor content of a test of critical thinking.

Related Research

Ennis (3) has considered the validity of the Watson-Glaser Critical Thinking Appraisal (hereafter referred to as the WG test) from a logical point of view. He asserts that the items in Parts 2, 3, and 4 of Form Bm of this test are keyed in such a way as to favor the pathological doubter. His reasoning may be illustrated by referring to Part 2 of the test, which is entitled "Recognition of Assumptions" and includes items 21 to 36 inclusive. For each item, the student is asked to decide whether a given statement "necessarily" assumes another given statement. Of the sixteen items in this part, only four are keyed "assumption necessarily made". Thus, the pathological doubter gets three-quarters of them right. Ennis then analyzes Parts 3 and 4 in the same way and shows that in the whole test the consistent pathological doubter could get up to sixty-seven out of the total of ninety-nine items correct. He, therefore, concludes that the keying so favors the pathological doubter that it is possible for him to be placed at the eighty-sixth percentile if he merely always chooses the "doubting" answer.

Frederiksen and Messick (4) have written an article in which they describe how they attempted to score set and content aspects of a test situation separately. Three tests of the report-writing type were used. The three hypotheses tested may be stated briefly as follows: a) set scores from all three tests measure the same thing reliably; b) set scores are related to personality or other set measures; and c) content scores measure the same things reliably. The investigators also tried to discover whether it was "criticalness" that the set scores measured, and whether verbal knowledge and reasoning were included among the things measured by the content scores. Frederiksen and Messick believe that if the answer to a question is not affected by knowledge, it is as if a subject merely tossed a loaded coin. They, therefore, assumed dichotomous data and derived a simple formula for set scores and another for content scores. They obtained correlation coefficients of 0. 82 when they intercorrelated content scores, 0.96 when they intercorrelated set scores, and 0. 05 when they intercorrelated content and set scores. Thus they were successful in separating content scores from set scores.

The Illinois Curriculum Program Committee (12) carried out a Critical Thinking Project in

*The author is deeply indebted to Dr. Kenneth B. Henderson and Dr. Philip J. Runkel of the University of Illinois for their assistance with the preparation of this article.

The computations for this paper were done on Illiac, an electronic computer of the Digital Computer Laboratory of the University of Illinois.

which a deliberate attempt was made to improve the critical thinking skills of a group of high school students. In order to measure differences between experimental and control groups, three tests were used: Form Bm of the WG test, A Test of Critical Thinking, Form G, prepared by the American Council on Education (hereinafter called the ACECT test), and A Test on Principles of Critical Thinking, Form F1.5 (hereinafter called the PCT test), prepared by the Illinois Critical Thinking Project Committee. The major hypothesis that knowledge of the principles of critical thinking is related to the ability to evaluate arguments and assess conclusions drawn was not confirmed. Moreover, low intercorrelations were obtained among the three critical thinking test scores. The Committee, then, began to think that the explanation for this might be in the measuring instruments themselves.

The author (9, 10), therefore, factor analyzed the item scores for these three tests. She used a sample of 587 students consisting of all students who had attempted all items on all three tests, and showed that this method of selecting the sample did not create bias. Phi correlation coefficients were computed and the centroid method of factoring employed. Factor analyses of the three critical thinking tests yielded only one weak general factor each. In view of the fact that intelligence tests, tests of psychomotor and physical abilities, performance tests, and tests of occupational abilities have all yielded a thoroughly established general factor (within a given test), the finding in this case was rather surprising. Moreover, several investigators who have factor analyzed reasoning tests using procedures similar to that used by the author, have successfully isolated several factors (1, 2, 5, 6, 8). The author suggested several explanations for her finding, one of which was that the population lacked randomness because of the presence of too many pathological doubters. It was thought possible that instruction in the principles of critical thinking had increased the subjects' tendency to be extremely cautious. The present study was, therefore, undertaken in the hope of learning more about the trait herein called pathological doubting.

Method and Results

Item scores on the three critical thinking tests mentioned above for the same sample of 587 students were used. An attempt was made to isolate those subjects suspected of being pathological doubters. Ennis' assertion that the items in subtests 2, 3, and 4 of Form Bm of the WG test are keyed to favor the pathological doubter led the author to use the scores of the subjects on these items as the basis for dividing the 587 students into two groups, namely, pathological doubters and non-doubters. Of the 65 items in subtests 2, 3, and 4, 52 of them are keyed as "assumption not necessarily made", "conclusion does not necessarily follow", or "conclusion does not follow beyond a reasonable doubt". The author arbitrarily decided to accept getting more than 33 of these items correct as evidence of pathological doubting. The criteria used in selecting pathological doubters, therefore, were: a) having at least 34 of the 52 items keyed "assumption not made" or "conclusion does not follow" correct, and b) having no more than 19 errors among the 65 items. The second criterion was added because it was thought that students who were making many errors, and especially if they were making errors of opposite kinds, were not consistent in their thinking and, therefore, probably were confused and perhaps just guessing the answers. These criteria yielded a group of 262 pathological doubters, and a group of 325 non-doubters.

Item scores for these subjects were available for Form Bm of the WG test, Form G of the ACECT test and Form F1.5 of the PCT test. The 204 items of these three tests were divided into 21 subtests. The items of the WG test have been grouped by the test makers into five parts: Inference (20 items), Recognition of Assumptions (16 items), Deduction (25 items), Interpretation (24 items), and Evaluation of Arguments (14 items). The 52 items of the ACECT test and the 53 items of the PCT test were grouped on the basis of natural breaks into nine and seven subtests respectively. Natural breaks sometimes were in the form of asterisks placed there by the test makers, and sometimes were merely considered "natural" by the author because they came between groups of similar items or items relating to one situation. These sixteen subtests, which varied in size from 3 to 13 items, were given the short appropriate labels shown in Table IV for easy reference. Subtest scores for the groups of pathological doubters and non-doubters were obtained.

Testing Hypothesis 1. — Pathological doubters chosen using criteria based on their responses to selected items of the WG test consistently exhibit a pattern of reasoning which differs from that of other subjects in the sample.

Mean subtest scores were computed for the 262 pathological doubters and for the 325 non-doubters. The difference between the mean subtest score for the group of pathological doubters and that for the group of non-doubters, for each of the 21 subtests, was tested for significance. In each case the variance was pooled and a t-test was made using this pooled variance. Wherever the frequency distribution was not approximately normal, a χ^2 test was also made.

For each of the 21 subtests, except subtest 20, a t-test indicated that the difference between the mean subtest score for the 262 pathological doubters and that for the 325 non-doubters was significantly different from zero beyond the .01 level of

significance. Differences of means were found to be significantly different from zero at the .01 level for nine of the ten χ^2 tests made. In the case of subtest 20 again, the difference was not significant. Table I summarizes these findings.

A second method was used to test this hypothesis. Answers to items 1 to 20 in subtest 1 of the WG test were scaled according to whether they gave positive or negative evidence of doubting. The scaling of these responses and the subsequent analysis were undertaken on the reasoning that people who give the doubting answer to some items (for example, those in subtests 2, 3, or 4) will give doubting answers to these items also. In subtest 1, which is called "Inference" by the test makers, a subject is required to label a statement as definitely true (T), probably true (PT), having insufficient data (ID), probably false (PF), or definitely false (F). The scale value allotted a subject for an item depended upon the keyed answer and the response given. Table II indicates how the various answers were scaled. This procedure may be illustrated by reference to item 2, which is keyed as probably true, and to a three point scale consisting of the points T or F, PT or PF, and ID. A person who is prone to accept statements as true without critically analyzing them might respond true. On the three point scale, this is one place to the left of the keyed response and indicates ready acceptance of a statement rather than a tendency to doubt. The answer, therefore, is given a scale value of -1. A response of ID is one place to the right of PT and indicates a tendency to be even more doubtful than the test makers consider warranted. Such a response was allotted a "doubting" value of +1. Other scale values were determined in a similar way. For each subject a score was allotted for each of the 20 items of subtest 1. An overall doubting index was thus obtained for each subject by adding the 20 scores. The mean doubting score for the group of 262 pathological doubters was compared with that for the group of 325 non-doubters, using a t-test.

The scaling procedure outlined in the preceding paragraph was repeated, except that in obtaining a doubting score for an item four additional scale values were also included. These are listed in the last four lines of Table II. The thinking behind this revision was that the direction of error does not matter; all that is of interest here is whether or not a subject chooses a "doubting" answer. Again, a t-test was used to test the significance of the difference of mean doubting scores of the two groups.

When the mean doubting index for the group of 262 pathological doubters (0.50) was compared with that for the group of 325 non-doubters (-0.31), a t value of 2.89 was obtained. When the revised scaling method was used, the difference between the mean doubting indices for the groups of 262 and 325 subjects (0.64 and -0.09 respectively) yielded a t value of 2.46. Thus the mean difference is significantly greater than zero beyond the .01 level of significance in the first case, and beyond the .02 level of significance in the second case.

Testing Hypothesis 2.—The presence of a response set such as pathological doubting affects the apparent factor content of a test of critical thinking.

A factor analysis of item scores for the WG test was carried out for the group of 325 non-doubters. Since the data were dichotomous, inter-item phi correlation coefficients were computed. The centroid method of factoring the correlation matrix was employed, and ones were placed in the diagonal cells of the matrix. The findings were compared with the results obtained by the author (9, 10) for the original sample of 587 which consisted of all subjects involved in the Illinois Critical Thinking Project who took all three critical thinking tests and who attempted all items on all three tests. Item scores for the PCT test were also factor analyzed using the same procedure.

As was the case in the author's original research (9, 10), no strong general factor was obtained for either the WG or the PCT test, when item scores for the 325 non-doubters were factor analyzed. In the case of the WG test the first centroid factor accounted for only 4.90 percent of the variance. For the PCT test this figure was 5.48 percent. For this reason, item scores for the ACECT test for the group of 325 non-doubters were not factor analyzed.

A factor analysis of subtest scores for the 21 subtests of the three critical thinking tests was carried out for the original sample of 587 subjects. Similar factor analyses were made for the group of 262 pathological doubters and the group of 325 non-doubters. Pearson product-moment correlation coefficients were computed. Unities were placed in the diagonal cells of the correlation matrix. Centroid factors were obtained and an attempt was made to rotate these factors to oblique simple structure using the oblimax procedure.

These factor analyses of subtest scores yielded one rather strong general centroid factor for each of the three samples. Here the first centroid factor accounted for approximately 32 percent of the variance in the case of the total sample of 587 subjects, 21 percent in the case of the 262 pathological doubters, and 22 percent in the case of the 325 non-doubters. Table III shows the variance accounted for by the first four centroid factors in the case of all three samples. Use of Guttman's "lower bound for the number of common factors" (7) indicated that an attempt should be made to rotate four factors to oblique simple structure in each case. The oblimax method of rotation was employed. This yielded quite high factor loadings, that is, high for reasoning factors. However, the

TABLE I

MEAN SUBTEST SCORES FOR THREE TESTS OF CRITICAL THINKING

Test and Subtest	Mean for 262 Pathological Doubters	Mean for 325 Non-Doubters	t*	χ^2*
Watson-Glaser				
1	12.81	10.49	6.26	
2	12.22	9.70	16.16	
3	21.45	14.99	7.87	230.00(1)**
4	17.52	14.02	8.18	
5	10.20	8.22	8.28	
ACECT				
6	5.54	4.74	7.15	154.17(3)
7	2.33	1.41	5.38	87.52(2)
8	4.37	3.09	13.66	
9	3.47	2.64	9.49	
10	8.13	6.66	7.69	50.79(2)
11	2.46	1.78	8.87	51.92(1)
12	2.20	1.33	10.33	68.25(1)
13	7.60	6.22	9.99	46.24(2)
14	4.26	3.29	6.36	45.04(3)
Principles				
15	4.38	3.75	6.70	21.57(2)
16	4.54	3.97	5.08	
17	3.42	2.79	5.27	
18	4.45	3.59	7.07	
19	7.06	5.52	9.07	
20	.89	.79	1.36	7.08(3)
21	5.41	3.98	9.77	

* All values of t and χ^2 are significant beyond the one percent level, except those for subtest 20.
**χ^2 values were computed where the distributions were not normal. Numbers of degrees of freedom are given in parentheses.

TABLE II

SCALE FOR OBTAINING A DOUBTING SCORE FOR ITEMS 1 TO 20 OF THE WATSON-GLASER TEST

Keyed Response*	Items	Evidence of Doubting: Positive — Answer Given	Positive — Scale Value	Negative — Answer Given	Negative — Scale Value
Original Method of Scaling:					
T	1, 17	PT	1		
T	1, 17	ID	2		
PT	2, 9, 16	ID	1	T	-1
ID	3, 5, 7, 8, 11, 12, 14, 18			F	-2
ID	3, 5, 7, 8, 11, 12, 14, 18			PF	-1
ID	3, 5, 7, 8, 11, 12, 14, 18			PT	-1
ID	3, 5, 7, 8, 11, 12, 14, 18			T	-2
F	4, 6, 13, 19	PF	1		
F	4, 6, 13, 19	ID	2		
PF	10, 15, 20	ID	1	F	-1
Revised Method of Scaling Contained Additions:					
T	1, 17	PF	1		
F	4, 6, 13, 19	PT	1		
PT	2, 9, 16			F	-1
PF	10, 15, 20			T	-1

*The abbreviation T stands for definitely true, PT for probably true, ID for insufficient data, PF for probably false, and F for definitely false.

TABLE III

THE VARIANCE ACCOUNTED FOR BY THE FIRST CENTROID FACTORS WHEN THE 21 SUBTEST SCORES ARE FACTOR ANALYZED

Centroid Factor	Percent of Variance Accounted For		
	Total Sample N = 587	Pathological Doubters N = 262	Non-Doubters N = 325
I	31.53	21.14	22.13
II	5.48	6.15	6.00
III	4.55	5.74	5.48
IV	4.36	5.27	4.86

oblimax solutions for the three samples were different. They are shown in Table IV.

Conclusions

The findings show the performance of the pathological doubters to be definitely and consistently superior to that of the non-doubters. They also show that those subjects who tend to choose a doubting answer in subtests 2, 3, and 4, also exhibit the same tendency when responding to items in subtest 1. There is, therefore, sufficient evidence to state that the pathological doubters do consistently exhibit a pattern of reasoning different to that of the other subjects in the original sample of 587, and thus to accept Hypothesis 1. One can also conclude that the doubters were successfully separated from the non-doubters using the author's arbitrarily chosen criteria which were based on Ennis' comments regarding the WG test.

It is particularly interesting to note that the pathological doubters consistently made higher scores than the other subjects on all subtests. Perhaps subtests 2, 3, and 4 in the WG test are not the only ones keyed to favor the pathological doubter. On the other hand, perhaps the pathological doubter's tendency to use extreme caution about accepting statements indicates that he can evaluate arguments, knows the rules of logic, recognizes unstated assumptions, and can invent and evaluate hypotheses, which are some of the skills the three critical thinking tests were designed to measure.

Subtest 20 in the PCT test seems to differ from the other subtests. It consists of three items only, and its mean item difficulty when computed for the sample of 587 was found to be only 0.28. These facts may account for the difference.

Seventy-two percent of the original sample of 587 students were in the experimental group. Seventy-nine percent of the group of 262 pathological doubters were from this group and only 66 percent of the non-doubters. Thus, there is some evidence that instruction in the principles of critical thinking increases the tendency to be overly cautious and to choose a "doubting" answer.

When item scores were factor analyzed, only slight differences were observable between the group of non-doubters and the whole sample of 587. No strong general factor was obtained in either case for either the WG or the PCT test. In contrast, when subtest scores were factor analyzed, four factors were obtained, perhaps because subtest scores are more significant and reliable than item scores. Differences in factor content, however, were apparent. Analysis of the factor loadings given in Table IV shows that the factors obtained for the three samples (the 587, the 262 pathological doubters, and the 325 non-doubters) are not sufficiently alike to be considered the same. Not even one factor common to the three samples has been found. There is, therefore, evidence here to support the hypothesis that the presence of a response set such as pathological doubting affects the apparent factor content of a test of critical thinking.

REFERENCES

1. Blotzum, W. A. "A Factorial Study of the Reasoning and Closure Factors," Psychometrika, XVI (1951), pp. 361-86.

TABLE IV

OBLIMAX FACTOR LOADINGS FOR SUBTESTS OF THREE TESTS OF CRITICAL THINKING

Subtest*	Items	No. of Items	Total Sample N = 587				Pathological Doubters N = 262				Non-Doubters N = 325			
			A**	B	C	D	E	F	G	H	I	J	K	L
1. Inference	1-20	20	-11	42	-26	37	101	-04	43	-25	-06	25	11	-38
2. Assumptions	21-36	16	15	79	-03	-04	64	25	09	-31	-41	15	21	22
3. Deduction	37-61	25	03	79	01	06	97	-09	37	-15	15	-13	56	01
4. Interpretation	62-85	24	02	77	-04	-10	02	57	33	21	32	63	60	-09
5. Arguments	86-99	14	00	-14	-03	72	24	-35	-11	24	-29	28	55	08
6. Direction of Evid.	1- 6	6	04	-03	02	70	-51	28	-19	67	19	-10	57	05
7. Inferences 1	7-10	4	05	76	-04	-08	70	20	80	-01	09	05	-51	99
8. Inferences 2	11-16	6	04	71	-02	05	86	01	42	-05	12	-17	57	-08
9. Generalizations	17-21	5	01	-05	01	72	-06	02	-33	43	-04	21	101	-38
10. Assumptions	22-31	10	00	07	08	65	-01	-02	31	69	-07	-07	27	43
11. Inferences 3	32-34	3	33	35	28	40	43	15	66	22	-15	-26	-02	53
12. Inferences 4	35-37	3	-03	16	46	21	38	-16	02	36	-06	-18	02	50
13. Hypotheses 1	38-46	9	-28	24	-27	40	30	-18	04	36	01	24	91	-20
14. Hypotheses 2	47-52	6	-66	03	-12	05	-07	-03	32	69	05	58	17	49
15. Definitions	1- 6	6	-36	-34	30	40	-33	24	-07	73	23	16	-34	84
16. Semantics 1	7-13	7	-75	-13	-13	08	-03	-06	29	63	37	59	66	-05
17. Fallacies	14-20	7	-73	-12	08	-11	-05	-01	-35	39	55	19	04	35
18. Semantics 2	21-27	7	-02	-02	59	16	03	04	-28	42	18	06	-37	81
19. Reasoning Rules	28-40	13	-37	38	21	-26	49	23	-05	03	46	14	08	32
20. Assumptions	41-43	3	11	-11	67	-11	-05	63	12	09	11	-48	15	-43
21. Inferences	44-53	10	-32	45	20	-24	15	45	27	32	46	19	18	32

* Subtests 1 to 5 are in the Watson-Glaser Test, subtests 6 to 14 in the ACECT test and subtests 15 to 21 in the Principles test.

**There should be two decimal places in all factor loadings. Decimal points have been omitted to save space.

2. Corter, H. M. "Factor Analysis of Some Reasoning Tests," Psychological Monographs, LXVI, No. 8 (1952), pp. 1-31.
3. Ennis, R. H. "An Appraisal of the Watson-Glaser Critical Thinking Appraisal," Journal of Educational Research, LII (1958), pp. 155-58.
4. Frederiksen, N., and Messick, S. "Response Set as a Measure of Personality," Educational and Psychological Measurements, XIX (1959), pp. 137-58.
5. Green, R. F., and others. "A Factor-Analytic Study of Reasoning Abilities," Psychometrika, XVIII (1953), pp. 135-60.
6. Guilford, J. P., and others. "A Factor-Analytic Study of Navy Reasoning Tests With the Air Force Aircrew Classification Battery," Educational and Psychological Measurements, XIV (1954), pp. 301-25.
7. Guttman, L. "Some Necessary Conditions for Common-Factor Analysis," Psychometrika, XIX (1954), pp. 149-61.
8. Matin, L., and Adkins, Dorothy C. "A Second-order Factor Analysis of Reasoning Abilities," Psychometrika, XIX (1954), pp. 71-78.
9. Rust, Velma I. Factor Analyses of Three Tests of Critical Thinking, unpublished Ph. D. dissertation, University of Illinois, 1959.
10. Rust, Velma I. "Factor Analyses of Three Tests of Critical Thinking," Journal of Experimental Education, XXIX (December 1960), pp. 177-82.
11. Rust, Velma I., and others. "A Factor-Analytic Study of Critical Thinking," Journal of Educational Research, in press.
12. Smith, B. O., and others. Teaching of Critical Thinking, to be published by University of Illinois Press.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 4, June 1961)

GEOMETRIC AID IN TEACHING RECTILINEAR CORRELATION

JOHN GAITO
Lake Forest College
Lake Forest, Illinois

THE PURPOSE of this paper is to present a simple geometric aid for use in the teaching of rectilinear correlation.

The coefficient of correlation may be concerned with the relationship between a number of variables. In the materials to follow, the concern will be with just two variables, X and Y. A convenient means of indicating the concept of correlation is by a diagram in which the X variable is plotted on the horizontal axis and the Y variable on the vertical axis. If the points within the axes scatter in an upward direction from left to right, in a manner which indicates that X and Y tend to vary together, a positive correlation is present. If the scatter is downward from left to right, a negative correlation is present, indicating that X and Y tend to vary inversely. If there appears to be no trend present, then zero correlation is indicated.

A convenient way of teaching rectilinear correlation is to consider the scattering present as describing an ellipse. An ellipse is completely defined by its major and minor axes. The major axis is the long axis, and the minor axis is the short axis. An ellipse varies in shape with two interesting limiting cases. One limit occurs when the major and minor axes are equal; then the figure is a circle. The other extreme is when the minor axis is equal to zero and the major axis is finite, in which case we have a straight line.

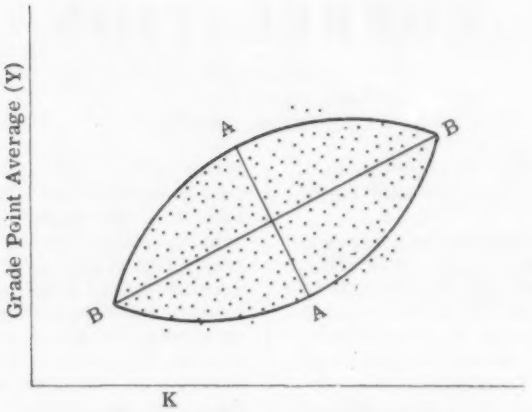
First, plot the points on the correlation chart, or scatter diagram. Thus one obtains an approximate elliptical figure by drawing a curve which encloses most points and passes through some of the outer points. Next draw a single straight line through the long axis of the figure. This line represents the trend of the scattering points. For descriptive purposes this can be handled quickly by inspection of the scatter and crudely drawing a line by hand. (A more precise method would be to draw the lines representing the regression of Y on X and X on Y, and then to use a line which is intermediate between the two to represent the trend.) This line would be the major axis of the ellipse. The minor axis would be a line drawn perpendicular to this line at its center.

Having drawn the scattering of the points to represent an ellipse, we then determine the magnitude of the relationship. This magnitude could be determined exactly by an appropriate formula available. The approximate magnitude, however, could be ascertained by looking at the ratio of the minor axis to the major axis, provided the two axes are drawn to the same scale. As the minor axis goes to zero (or the major axis gets very large relative to the minor axis), this ratio will also go to zero, and the correlation coefficient will vary toward one. When the minor axis equals zero and the major axis is finite, the ratio also equals zero, the coefficient is one, no spread is present, and we have a single straight line—with perfect correlation. (An exception occurs when the slope of the line representing the major axis equals zero. In this case, the Y variable equals a constant value for every X, and the line is parallel to the X axis. Thus, there is no change in Y with a change in X; Y remains the same. In this case, even though the minor axis is zero and the ratio is zero, the correlation coefficient would be zero because no relationship is present.) On the other hand, as the minor axis increases, with the major axis relatively fixed, the ratio increases and the coefficient decreases. When the minor axis is equal to the major axis, the ratio is one, the coefficient is zero, the spread is maximal, and we have a circle with no relationship present.

This geometrical interpretation has value beyond the elementary teaching of correlation. For example, it is found useful in looking at the effects of range of ability in the sample on the size of the coefficient. Let us use an example in which 500 Ss have scores on a college entrance examination, all are accepted, grade-point averages are available later for each, and the relationship between examination scores and grade-point average is .60. If one were to plot these data on a correlation chart it would be found that a ratio of approximately 2/5 would occur. The ratio might be greater or less, but the magnitude of the ratio is not important for this example. Only the relative aspects are of consequence. Thus, let us assume that 2/5 is the ratio. This situation is shown in Figure 1. Now if, instead of allowing all 500 applicants to enter school, we had used a cutting score on the examination, the diagram would be

FIGURE 1

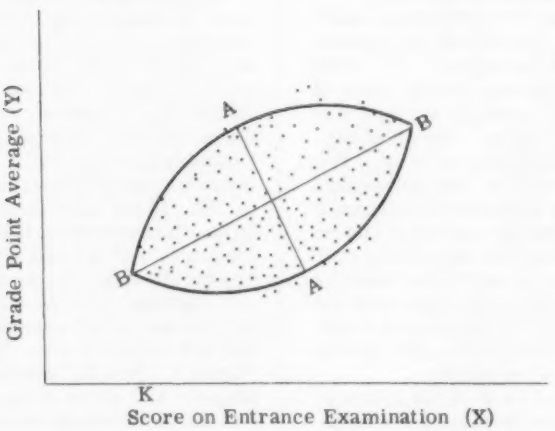
A HYPOTHETICAL EXAMPLE IN WHICH ALL APPLICANTS ARE ACCEPTED AT A COLLEGE



Note: The relationship indicated is .60. The ratio of the minor axis (A) to the major axis (B) is assumed to be 2/5. Point K represents the cutting score which is used in Figure 2.

FIGURE 2

THE RELATIONSHIP INDICATED WHEN A NUMBER OF APPLICANTS TOWARD THE LOWER END OF THE X VARIABLE, BELOW K, HAVE BEEN ELIMINATED



changed as in Figure 2. Note that the minor axis has been shifted from its position in Figure 1 so as to be perpendicular to the major axis at the latter's center; however, the minor axis is approximately the same in magnitude as it was in Figure 1. The major axis has decreased in size so that the ratio now is approximately 2/4. This increase in ratio would indicate a lesser degree of relationship than is present in Figure 1. Such a procedure easily illustrates the effect of composition of the sample (as indicated by the range of the X variable) on the magnitude of the relationship, with a greater degree of relationship present when the sample is heterogeneous than when the sample is homogeneous.

We have not developed an exact mathematical relationship between the size of the ratio and the size of the correlation coefficient. Intuitively, however, we assume that there is such a relationship, so that one could indicate the size of the coefficient by knowing the size of the ratio. This geometrical interpretation of correlation, using an elliptical ratio, should be useful for teaching purposes even if it is merely of an ordinal nature.